

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»**

**Інститут прикладного системного аналізу
Кафедра математичних методів системного аналізу**

«На правах рукопису»
УДК 519.248

«До захисту допущено»
Завідувач кафедри
_____ О. Л. Тимошук
«__» _____ 20__ р.

**Магістерська дисертація
на здобуття ступеня магістра
зі спеціальності 124 Системний аналіз
на тему: «Система оцінки ефективності фінансових витрат у
виробництві з використанням системного підходу»**

Виконала:
студентка II курсу, групи КА-71мп
Козлова Марія Олександрівна _____

Керівник:
к.т.н., доцент кафедри ММСА
Кузнєцова Н.В. _____

Рецензент:
завідувач кафедри загально-інженерних дисциплін та теплоенергетики
Таврійського національного університету
д. т. н., професор Медведєв М. Г. _____

Засвідчую, що у цій магістерській
дисертації немає запозичень з праць
інших авторів без відповідних
посилань.
Студентка _____

Київ
2018

**Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»**

Інститут прикладного системного аналізу

Кафедра математичних методів системного аналізу

Рівень вищої освіти – другий (магістерський)

Спеціальність (спеціалізація) – 124 «Системний аналіз» («Системний аналіз і управління»)

ЗАТВЕРДЖУЮ

Завідувач кафедри

_____ О.Л. Тимошук

«__» _____ 20__ р.

**ЗАВДАННЯ
на магістерську дисертацію студенту**

Козловій Марії Олександрівні

1. **Тема дисертації** «Система оцінки ефективності фінансових витрат у виробництві з використанням системного підходу», науковий керівник дисертації Кузнєцова Наталія Володимирівна, к.т.н. доцент кафедри ММСА, затверджені наказом по університету від «07» листопада 2018 р. №4112-с

2. **Термін подання студентом дисертації** _____

3. **Об'єкт дослідження:** статистичні дані щодо фінансово-виробничих процесів

4. **Предмет дослідження:** математичні моделі і методи дисперсійного та коваріаційного аналізу, методи логістичної регресії

5. Перелік завдань, які потрібно розробити:

- 1) Огляд технічної літератури за темою роботи
- 2) Дослідження актуальності обраної теми
- 3) Вибір методів для моделювання і прогнозування;
- 4) Збір вхідних даних;
- 5) Виконання обчислювальних експериментів;

- 6) Аналіз результатів моделювання і прогнозування;
- 7) Проведення аналізу ринкових можливостей запуску стартап-проекту;
- 8) Підготовка ілюстративного матеріалу;
- 9) Оформлення пояснювальної записки

6. Орієнтовний перелік графічного (ілюстративного) матеріалу

- 1) Постановка завдання дослідження
- 2) Методи обробки і аналізу зображень
- 3) Наукова новизна результатів

7. Орієнтовний перелік публікацій:

Система оцінки ефективності фінансових витрат у виробництві з використанням системного підходу // Системні дослідження та інформаційні технології. – Стаття подана в редакцію журналу

8. Дата видачі завдання _____

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації	Примітка
1.	Отримання завдання на дипломну роботу	07.09.2018 – 09.09.2018	
2.	Огляд технічної літератури за темою	10.09.2018 – 30.09.2018	
3.	Дослідження актуальності обраної теми	01.10.2018 – 07.10.2018	
4.	Вибір методів для моделювання і прогнозування	08.10.2018 – 14.10.2018	
5.	Збір вхідних даних	15.10.2018 – 21.10.2018	
6.	Виконання обчислювальних експериментів	22.10.2018 – 28.10.2018	
7.	Аналіз результатів моделювання і прогнозування	29.10.2018 – 04.11.2018	
8.	Проведення аналізу ринкових можливостей запуску стартап-проекту	05.11.2018 – 11.11.2018	
9.	Підготовка ілюстративного матеріалу	12.11.2018 – 18.11.2018	
10.	Оформлення пояснювальної записки	19.11.2018 – 26.11.2018	

Студент

М.О. Козлова

Науковий керівник дисертації

Н.В. Кузнєцова

РЕФЕРАТ

Магістерська дисертація: 91 с., 29 рис., 24 табл., 2 додатки і 19 джерел.

Об'єкт дослідження – виробничі лінії на виробництві, витрати які виникають у процесі виробництва.

Предмет дослідження – моделі регресійного та дисперсійного аналізів та методи інтелектуального аналізу даних для кількісного та якісного оцінювання показників.

Мета роботи – розробка моделей та методів моделювання розподілення фінансових витрат та їх порівняння із існуючими загальноприйнятими.

Методи дослідження – моделі регресійного аналізу, дисперсійного аналізу, побудова байесовських мереж.

У цій роботі наведені результати побудови моделей аналізу взаємозв'язків різних факторів виробництва. Проведено порівняльний аналіз отриманих моделей за допомогою різних критеріїв, а також зроблено висновки щодо їхньої точності. Виявлено, які саме причини є важливими, запропоновані фінансово ефективні дії для мінімізації витрат. Обрано найкращі моделі для побудови залежностей. Тому для подальших досліджень рекомендовано використовувати саме такі моделі.

За матеріалами магістерської дисертації були написані тези та наукова стаття. Стаття опублікована в електронній збірці доповідей у видавництві Інтернаука.

Прогнозні припущення щодо подальшого розвитку об'єкта дослідження – вдосконалення існуючих моделей, знаходження нових факторів виробництва та інше. А також покращення існуючої системи прийняття рішень на основі побудованих моделей.

ДИСПЕРСІЙНИЙ АНАЛІЗ, СТАТИСТИЧНИЙ АНАЛІЗ ДАНИХ, ВИРОБНИЧА ЛІНІЯ, МІНІМІЗАЦІЯ ВИТРАТ, КОВАРІАЦІЙНИЙ АНАЛІЗ, ЛОГІКО-ІМОВІРНІСНА МОДЕЛЬ.

ABSTRACT

Master's dissertation: 91 pp., 29 figures, 24 tables, 2 annexes and 19 sources.

The object of research - production lines in production, costs incurred in the production process.

The subject of the study is regression and dispersion analysis models and methods of data mining for quantitative and qualitative evaluation of indicators.

The purpose of the work is the development of models and methods for modeling the distribution of financial costs and their comparison with existing generally accepted.

Methods of research - models of regression analysis, dispersion analysis, construction of Bayesian networks.

In this work the results of construction of models of analysis of interconnections of various factors of production are presented. A comparative analysis of the obtained models was carried out with the help of various criteria, as well as conclusions were made regarding their accuracy. It is found out which causes are important, offered financially effective actions for minimization of expenses. The best models for constructing dependencies are chosen. Therefore, it is recommended to use such models for further research.

The thesis and the scientific article were written on the materials of the master's dissertation. The article was published in the electronic collection of reports at the International Science Academy publishing house.

Foreseeable assumptions about the further development of the research object - improvement of the existing model, finding new factors of production, and so on. As well as improving the existing decision making system based on the built models.

DISPERSION ANALYSIS, STATISTICAL ANALYSIS OF DATA, PRODUCTION LINE, MINIMIZATION OF EXPENSES, QUALIFICATION ANALYSIS, LOGIC-IMPORTANT MODEL.

ЗМІСТ

ЗМІСТ

УМОВНІ ПОЗНАЧЕННЯ ТА СКОРОЧЕННЯ	8
ВСТУП	9
РОЗДІЛ 1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ	12
1.2 Загальні відомості щодо предметної області	12
1.2 Аналіз об'єкту дипломної дисертації	14
1.3 Огляд існуючих підходів у рішенні схожих проблем на виробництвах	16
1.4 Постановка задачі дослідження	17
Висновки до розділу 1	18
РОЗДІЛ 2 Методика прогнозування	19
2.1 Методи дисперсійного аналізу	19
2.2 Методи коваріційного аналізу	33
2.2 Лінійна імовірнісна модель	40
2.4 Збір даних	41
2.5 Визначення та обробка пропусків	42
2.6 Оцінка кваліфікації співробітника	42
Висновки до розділу 2	51
РОЗДІЛ 3 СИСТЕМА ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ ВИЗНАЧЕННЯ ПРИЧИН ВИТРАТ НА ВИРОБНИЧІЙ ЛІНІЇ	52
3.1 Перевірка на нормальність розподілу зворотніх відходів	52
3.2 Побудова архітектури програми по підтримці прийняття рішень	62
3.3 Основні технічні вимоги для коректної роботи програми	65
РОЗДІЛ 4 СТАРТАП РОЗДІЛ	79
4.1 Опис ідеї проекту та її технологічний аудит	79
4.2 Аналіз ринкових можливостей запуску стартап проекту	80
4.3 Розробка ринкової стратегії проекту	85
4.4 Розробка маркетингової програми стартап проекту	87
4.5 Висновки	88
ВИСНОВКИ	89
Перелік посилань	91
Додатки	Ошибка! Закладка не определена.

УМОВНІ ПОЗНАЧЕННЯ ТА СКОРОЧЕННЯ

ЗВ – зворотні відходи

ANOVA – analysis of variance

ANCOVA – analysis of covariation

ДА – дисперсійний аналіз

ЛІМ – логіко-імовірнісна модель

КА – коваріаційний аналіз

СППР – система підтримки прийняття рішень

ВСТУП

На території України є велика кількість виробництв, які через економічний стан втратили розуміння за яким вектором далі вести свою діяльність. На сьогоднішній день динаміка економічної закритості України від Росії має дуже неоднозначні наслідки. З одного боку, підприємства внаслідок санкцій втратили такий великий регіон споживачів, як Росія, а це близько 140 мільйонів. А отже їх продажі помітно впали, також деякі виробництва втратили потужності розташовані на ворожнечій території.

На рисунку 1 продемонстровано приклад динаміки об'ємів виробництва з 2012 по 2018 однієї компанії. Можна побачити, як дійшовши ями об'єми тільки потрохи починають відновлюватись.

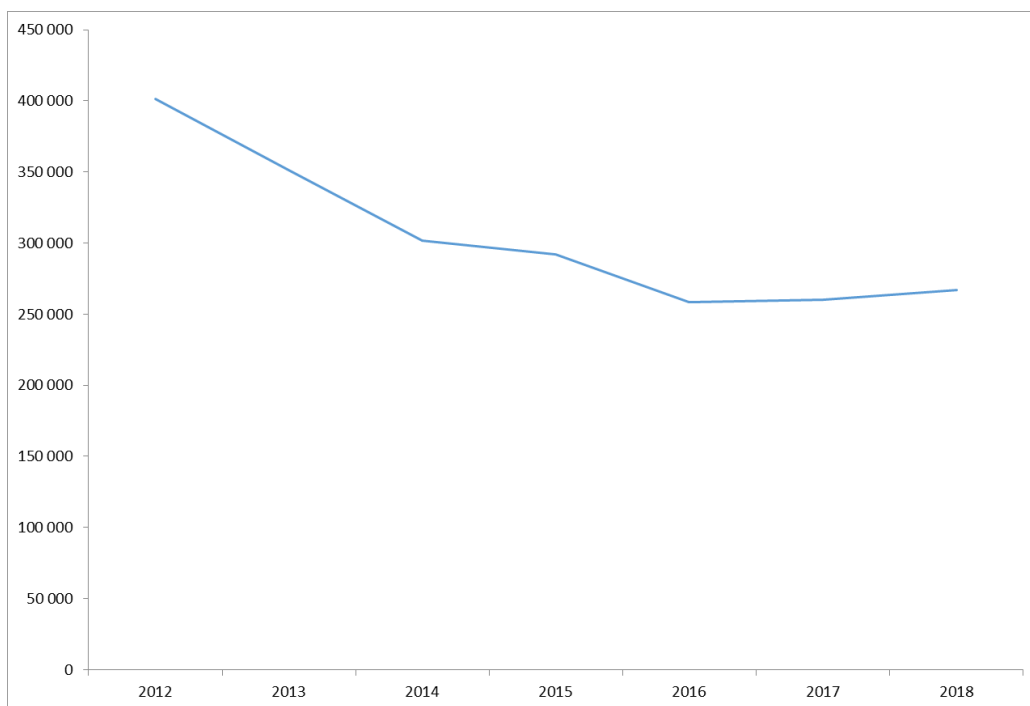


Рисунок 1. Динаміка об'ємів виробництва з 2012 по 2018 рік

Але з іншого боку, характер конкуренції змінився. Замість великої кількості дешевих низькоякісних товарів, на полицях з'являються більш якісні дорогі товари європейських виробників. Оскільки купівельна спроможність

українських споживачів поки що залишається на старому рівні, що не дає змоги купувати якісні товари, кількість конкурентів падає. Тобто це шанс якомога більше покрити сегмент українських споживачів. Паралельно з цим вкрай необхідно проводити зміни на виробництвах, раціоналізувати, покращувати та пристосовувати свої потужності, з тим щоб почати конкурувати на дійсно високому рівні вже з європейськими виробниками.

Це обумовлює *актуальність* проблеми. Оскільки основою великоприбуткових виробництв є виробничі лінії, то необхідно зосередитись на процесі виробництва. Слід розглянути лінію як систему, тоді можна побачити цілий ряд невизначеностей: якість етикетки, відмови устаткування, розподіл фінансування і одне з головного це трудові ресурси. Через економічну нестабільність ми бачимо великий відтік трудових ресурсів в інші країни

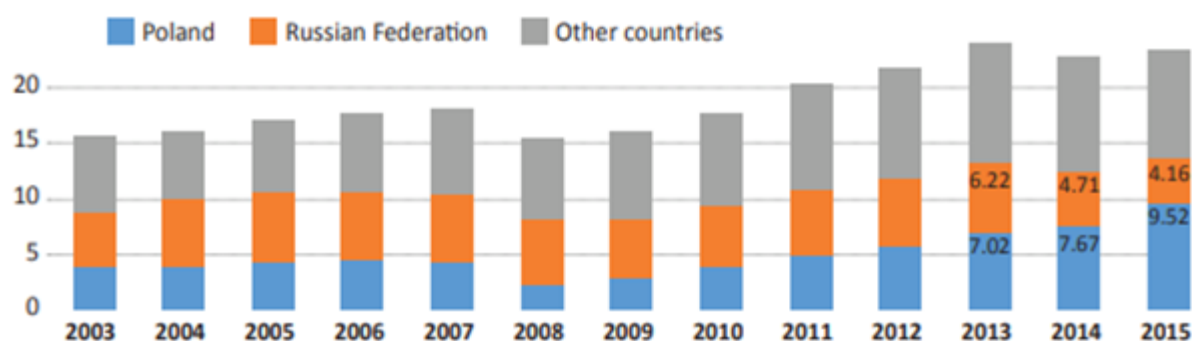


Рис. 2 Статистика перетинів кордону українцями, міл. разів

Через це ми маємо розуміння, що всі ці невизначинності потрібно мінімізовувати.

Таким чином, *мета дослідження* – це розробка моделей та методів моделювання розподілення фінансових витрат та їх порівняння із існуючими загальноприйнятими. Для її досягнення були поставлені наступні завдання:

- 1) Визначити головні причини виникнення втрат на виробничій лінії
- 2) Виявити найбільш актуальні та перспективні фінансові показники, які будуть використані у моделі

- 3) Визначити які фактори найбільше впливають на досліджуєму систему.
- 4) Розробити власні моделі прогнозування показників
- 5) Розробити прототип системи прийняття рішення для розв'язку схожих завдань на інших лініях у подальшому

Об'єктом цього дослідження є виробничі лінії на виробництві, та їх особливості, для ефективного планування фінансових показників цієї лінії, а *предметом* – моделі регресійного та дисперсійного аналізів та методи інтелектуального аналізу даних для кількісного та якісного оцінювання показників. При дослідженні серед інших застосовувалися наступні методи: аналіз, синтез, абстрагування, порівняння та узагальнення.

Практичним результатом роботи є розроблений програмний продукт, який дозволяє оцінювати фінансові показники на основі методів інтелектуального аналізу даних

РОЗДІЛ 1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.2 Загальні відомості щодо предметної області

На практиці не завжди потрібно визначати результати всіх ресурсів.

Багато підприємств визначають ефективність тих чинників, які вважаються придатними показниками діяльності й успіху підприємства. Отже, практичне визначення загальної ефективності завжди потрібно формулювати як зосередження на обмеженій кількості ресурсів, які в результаті вказують на успіх підприємства.

При ізольованому використанні індивідуальні показники можуть не створювати об'єктивної картини. Саме зниження ефективності використання 1-ого ресурсу може бути необхідним для підвищення ефективності використання іншого. Такий компроміс бажаний тоді, коли знижуються загальні витрати, але результат буде втрачено, якщо використовувати будь-який частковий показник окремо. На основі спостережень наслідків зростання загальної ефективності менеджерам можна робити висновки про загальний результат ефективності. Якщо в цілому все покращиться. Через можливість існування компромисного вибору загальний показник ефективності повинен оцінювати сукупні фінансові наслідки і, отже, має бути фінансовим показником.

Показники господарської діяльності підприємств, класифікуються за групами:

- 1) Норми
- 2) Показники часу роботи (затраченого на виробництво продукції)
- 3) Показники трудових ресурсів
- 4) Кількість зворотніх відходів
- 5) Показники виробництва
- 6) Фінансові показники

7) Якість виробів тощо.

Показники господарської діяльності підприємницьких структур часто можна об'єднати в певну систему (рис. 1.1).



Рисунок 1.1 – Структура системи показників

Тому потрібно застосовувати у підприємницькій діяльності різні показники щодо господарських операцій і процесів, це дає змогу всеоб'ємно відображати різні особливості діяльності підприємств, контролювати їх та визначати реальні всебічні результати цієї діяльності.

Кожний обраний показник діяльності підприємства звісно має об'єктивну основу й економічний зміст, він повинен відображати процес розширеного відтворення і його основні елементи: виробництво, витрати, розподіл, обмін і споживання. Діаграма найважливіших чинників, що визначають ефективність виробництва, продемонстровано на рис. 1.2. Кожний показник системи оцінки, узятий окремо, крім узагальнюючого показника, є лише частковою характеристикою ефективності.



Рисунок 1.2 – Найважливіші чинники, що визначають ефективність виробництва

1.2 Аналіз об'єкту дипломної дисертації

Що ми будемо аналізувати, та яку систему підтримки прийняття рішення ми будемо. На кондитерському виробництві основна діяльність це вироблення цукерок, як ми розуміємо при великих тонажах їх виготовляють не вручну, а великими, потужними лініями. Ці лінії потребують великої уваги як висококваліфікованим обслуговуючим персоналом, так і уваги менеджерів які визначають куди витрачати гроші: на модернізацію якогось вузла, чи на навчання персоналу чи інше.

Тому, виберемо одну лінію на прикладі якої ми продумаємо систему для прийняття рішення.

Для початку проаналізуємо які причини зупинок того чи іншого вузла можуть бути (для коректного проектування будь-якої іншої лінії):

- 1) Тип продукту (наприклад, продукти з вкрапленнями більш складні у виробництві)
- 2) Помилка обслуговуючого персоналу
- 3) Погана сировина чи етикет

За статистикою це основні причини зупинок, інші є не впливовими та їми можна знехтувати.

До чого призводять зупинки, та чому цьому явищу потрібно приділяти увагу:

- 1) Збільшення зворотніх відходів. Зворотні відходи є двох типів ті що можна переробити, та ті що ми повинні викинути. І той і той випадок впливає на наші витрати. Тому будемо рахувати витрати від зворотних відходів, як вартість сировини яка входить до цих зворотніх відходів.
- 2) Заробітня плата персоналу який чекав поки лінію запустять знов.
- 3) Час який не вироблялась продукція, тобто як недовипущена продукція = недоотриманий прибуток
- 4) Вірогідність, що працівник звільниться, і нам знову прийдеться витратити ресурси на його навчання

Які основі важелі впливу ми пропонуємо:

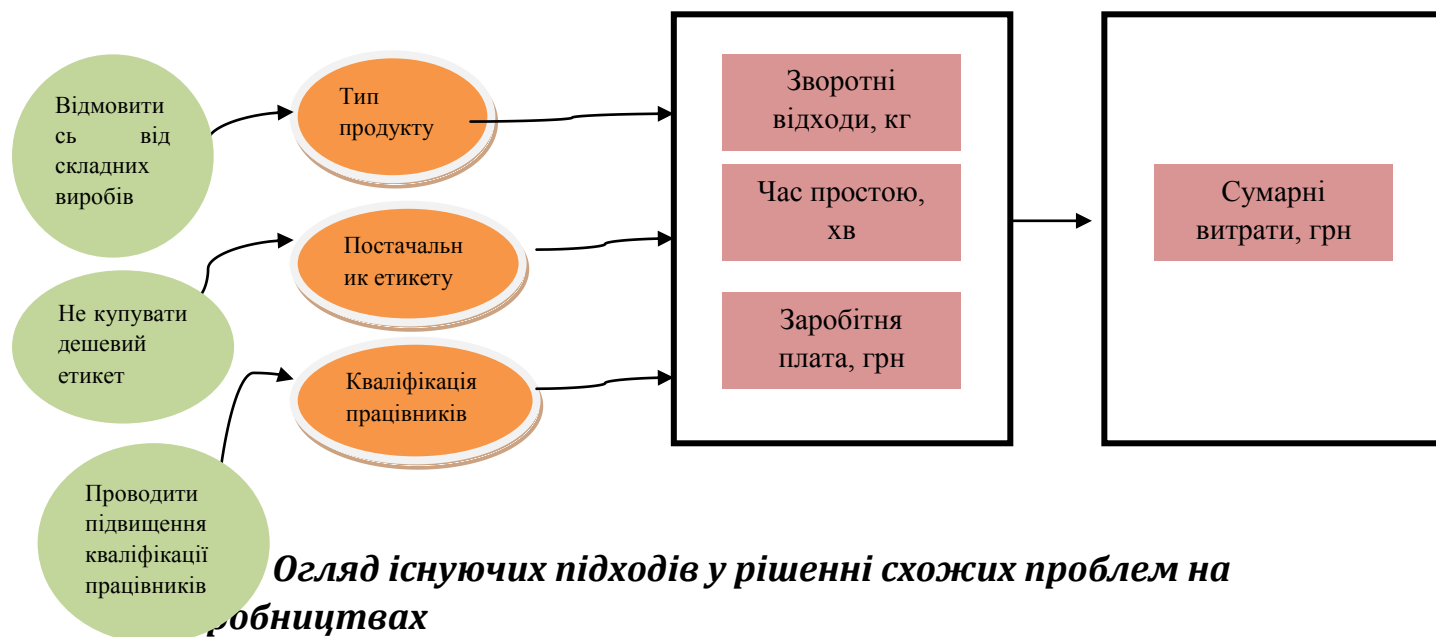
- 1) Збирати статистику по відмові устаткування, аналізуючи яку можна запобігти більшості раптових зупинок.
- 2) Навчати персонал, навчати працювати на різних лініях тим самим збільшувати досвід працюючих
- 3) Впроваджувати первинний контроль якості етикету чи сировини. Проаналізувати чи є залежність від ціни на етикет на якість етикету, якщо так то визначити що вигідніше: економити на етикеті, чи зменшити кількість зупинок і як наслідок зменшити зворотні відходи.

Ці кроки дозволять мінімізувати витрати.

Для того, щоб побудувати модель системи виробнича лінія визначимо фінансові показники які будуть показниками ефективності розподілення витрат:

- 1) Для того щоб прийняти рішення, чи виробляти складний продукт, який невід'ємно тягне за собою великі зворотні відходи, ми беремо маржинальність продукту
- 2) Для оцінки ефективності вкладання грошей у модернізацію чи навчання персоналу беремо окупованість.

На основі вищевикладених причин проблеми шляхів вирішення спробуємо зобразити це у вигляді діаграми.



Огляд існуючих підходів у рішенні схожих проблем на виробництвах

Як вже було сказано вище ми маємо три проблеми тобто три невизначеності. З кожною з цих невизначеностей великі виробництва поступають по різному. Наприклад, компанія, що робить Chupa-Chups, як можна помітити робить один максимум два види продукції. Це робиться для того щоб запобігти першій проблемі – складності продукту. Відточучи майстерність на одному-двох продуктах компанії мінімізують свої витрати тим самим захищають себе від неочікуваних затрат.

Друга проблема у європейських фірмах виникає рідше за українських, тому що їм дуже важливо виглядати якісно на полицях, тому їх рішення це не ризикувати на економії на етикеті, а просто обрати 1-2 постачальників та купляти стабільно дорогий, але якісний етикет.

Підвищення кваліфікації співробітників це дуже розповсюджений метод, фірми що піклуються про співробітників залюбки вкладають у них гроші через це відчують потужний фідбек. Ця методика діє не тільки у лінійному смислі, а ще починають діяти різні психологічні фактори. Але варто пам'ятати,

що завжди є ймовірність, що співробітник пройшовши навчання може піти з фірми. Тобто потрібно оцінити вірогідність, що співробітник піде з компанії, гроші будуть витрачені даремно.

1.4 Постановка задачі дослідження

Мета магістерської дисертації є дослідження та покращення, вдосконалення існуючих методів аналізу і прогнозування витрат на виробничій лінії, розробка програмного забезпечення для попередньої обробки даних і побудови моделей дисперсійного та коваріаційного аналізів, а також лінійно-імовірносних моделей для оцінки ймовірності втрати співробітника, та перевірка побудованої моделі на адекватність. У рамках дисертації було зроблено, вирішено наступні задачі:

- розроблено архітектуру системи підтримки прийняття рішень (СППР) для аналізу, моделювання та прогнозування залежності витрат від існуючих факторів;
- розробити програму для побудови лінійно-імовірносної моделі для оцінки додатков витрат;
- протестовано комп'ютерну програму на реальних даних з виробництва та проведено порівняльний аналіз з іншими методами.

Для вирішення цих задач необхідно було досліджувати вже існуючі інтелектуальні рішення для вирішення задач керування фінансовими ризиками. Об'єкт дослідження – статистичні дані щодо рівня зворотніх відходів, кількості зупинок та тривалості простоїв виробничої лінії, які потребують якісної, впевненої аналітичної обробки та є необхідними для побудови моделей та прийняття рішень при визначенні мінімумів витрат. Предмет дослідження – математичні методи побудови моделей оцінки, а саме: дисперсійний аналіз, коваріаційний аналіз, логіко-імовірносна модель.

Висновки до розділу 1

У першому розділі було сформульовано проблему дисертації. Було визначено основні причини та шляхи вирішення цих проблем. Далі розглянемо математичний апарат завдяки якому ми будемо вирішувати поставлені проблеми, та на основі якого будемо будувати систему підтримки прийняття рішення. Також у першому розділі було розглянуто як вирішують схожі проблеми іншу виробництва.

РОЗДІЛ 2 Методика прогнозування

2.1 Методи дисперсійного аналізу

Чітке розуміння математичної моделі ДА (дисперсійного аналізу) допомагає зрозуміти необхідні обчислювальні операції. А особливо воно є необхідним під час обробки даних багатфакторних дослідів, що поєднують значно більше джерел варіювання, аніж прості однофакторні досліді. Наприклад, в двофакторному досліді, поставленому методом звичайних повторень, сума квадратів для варіантів $C \times V$ розчленовується на 3, а в трьохфакторному - на 7 компонентів.

Загальна сума квадратів цих дослідів описується виразом:

$$C\bar{Y} = (CA + CB + CAB) + CZ$$

$$C\bar{Y} = (CA + CB + CC + CAB + CAC + CBC + CAB C) + CZ,$$

де в дужках вказані вирази – сума квадратів досліджуваних трьох факторів А, В та С.

Багатфакторний дисперсійний комплекс - це об'єднання всіх вихідних спостережень, які дозволяють оцінити дію та взаємозв'язок декількох факторів ґрунтуючись на статистичних даних, які досліджуються на змінність результативної ознаки.

Ефект взаємодії проставлять ту частину загального переменчивості, яка викликана різними діями 1-го фактора при різних градаціях іншого. Специфічна дія поєднань в експерименті виявляється при одній градації першого фактора а другий діє слабо, а при іншій градації він проявляється сильніше а й стимулює розвиток результативної ознаки.

При експериментальних дослідженнях можна помітити те, що ефект досліджуваних факторів зазвичай більший (ефект синергії) чи менший (антагонізм) аніж сума цих ефектів, застосованих порізно. Відповідно робимо висновок про існування взаємодії факторів . Якщо в першому разі воно носить позитивний характер, то в другому- кардинально протилежний. У випадку,

коли досліджувані фактори не взаємодіють, надбавка за спільне їх застосування рівне сумі приросту від окремого впливу. Інакше це явище називають аддитивізмом.

ДА даних багатфакторного комплексу складається з двох етапів.

Перший етап. Розкладається загальна варіація ознаки на варіювання варіантів та залишкове.

$$C_Y = C_V + C_Z.$$

Другий етап. Сумму квадратів відхилення для варіантів розкладають на компоненти, відповідно до джерел варіювання, а саме: головні ефекти досліджуваних факторів, їх взаємодія та наслідки.

У випадку двофакторного дослідження вираз прийме значення:

$$C_V = C_A + C_B + C_{AB},$$

А при трьофакторному:

$$C_V = C_A + C_B + C_C + C_{AB} + C_{AC} + C_{BC} + C_{ABC}.$$

ДА двофакторного аналізу вивчення градацій фактора А та градацій фактора В (враховуючи варіантів 1 А та В), що проводяться n разів, складаються з таких етапів:

Перший етап. Визначаємо суму та середні за кожним з варіантів, загальну суму та середній урожай виходячи з досвіду.

Другий етап. Обчислюються загальна сума квадратів відхилень, сума квадратів для варіантів та залишків

$$N = l_A \cdot l_B \cdot n;$$

$$C = \frac{(\sum X)^2}{N};$$

$$C_Y = \sum X^2 - C;$$

$$C_V = \sum \frac{V^2}{n} - C;$$

$$C_Z = C_Y - C_V;$$

Для обрахування сумм квадратів за такими чинниками А, В і взаємозв'язок дії АВ складається допоміжна таблиця, в яку записуються сумми за варіантами. Підсумовуючи цифри, знаходяться суми А, суми В і обчислюються суми квадратів відхилень для головних ефектів і взаємодії.

Запишемо суму квадратів для фактора А при (1 А-1) ступенях свободи:

$$C_A = \sum \frac{A^2}{l_B \cdot n} - C$$

Запишемо суму квадратів для фактора В при (В-1) ступенях свободи:

$$C_B = \sum \frac{B^2}{l_A \cdot n} - C$$

Суму квадратів взаємодії факторів А та В – різниця при (А-1) та (В-1) ступенях свободи

$$C_{AB} = C_V - C_A - C_B$$

У таблиці 1 продемонстровано приклад таблиці дисперсійного аналізу, до якої записують суму квадратів та визначають фактичні значення критерію F на її основі

Таблиця 1. - Результати двофакторного аналізу

Варіація	Сума квадратів	Число ступеней свободи	Дисперсія	Критерій Фішера, F _{факт}	Рівень значущості, р	Критерій Фішера, F _{табл}
Фактора А						
Фактора В						
Взаємодія						
Випадкова						
Загальна						

Доцільне використання ДА для обробки статистичного матеріалу передбачає однорідність дисперсій за вибірками, нормальне або близько до нього розподіл варіюють величин, значення кількості отримують незалежно одне від іншого. У дослідженнях незалежність порівняння досягається випадкуваннями розміщеним варіантів в досвіді і випадковим відбором проб до вибірки. Коли є підстави припускати неоднорідність дисперсій по вибіркам, про

що зазвичай свідчать великі відмінності в варіюється, то рекомендується трансформувати вихідні дані.

Найбільш підходящі і найчастіше застосовуються перетворення наступні:

- Логарифмічні - кожне значення X трансформується в $\lg X$ або в $\ln (X - 1)$, якщо деякі спостереження дорівнюють нулю;
- трансформація даних підрахунку чисельності шляхом знаходження квадратного кореня із X , т. е. \sqrt{X} або $\sqrt{X+1}$, коли декілька спостережень дають нульові або дуже невеликі значення.

Перетворені значення обробляються за схемою дисперсійного аналізу після проведених оцінок переходять назад до первинних одиницях виміру. Середні, отримані в процесах перетворення, будуть дещо виділятися від середніх, отриманих за вихідними даними, але різниця зазвичай не велика, і найбільш правильним середнім буде значення, отримане зворотним переходом.

Всі кінцеві результати дисперсійного аналізу можна отримати без обчислення дисперсій, на основі тільки сум квадратів.

У дисперсійному аналізі ортогональних комплексів застосовуються адитивні властивості часткових дисперсій – сума квадратів цент. відхилень, що записуються у вигляді:

$$C_V + C_Z = C_Y, \quad C_A + C_B + C_{AB} + C_Z = C_Y.$$

На властивості адитивності часткових дисперсій ґрунтується закон розкладання загальних дисперсій в ортогональних комплексах.

Запишемо відношення часткових дисперсій до загальної:

$$\frac{C_V}{C_Y} + \frac{C_Z}{C_Y} = \frac{C_Y}{C_Y} = 1 \quad \frac{C_A}{C_Y} + \frac{C_B}{C_Y} + \frac{C_{AB}}{C_Y} + \frac{C_Z}{C_Y} = \frac{C_Y}{C_Y} = 1$$

Кожне відношення показує долю участі окремої приватної дисперсії в навчанні загальної.

Відштовхуючись від того, що кожен приватну дисперсію спів ставляють відповідному приватному впливу, то ставленні приватної до загальної дисперсії

оцінює долю цього впливу на загальний статистичний вплив усіх чинників, які встановлюють розвиток даної результативної ознаки.

Відповідно, частка, що виражається у відносних одиницях чи відсотках, кожної приватної дисперсії у загальній сумі приймається як показник впливу, що характеризує дану приватну дисперсію – чи однією з факторіальних, чи випадковою.

Якщо взяти, наприклад, в однофакторном комплексі, то чим більшу частку в загальній дисперсії займає її факторіальна частина (C_V/C_Y), тим більше буде частина загальної різноманітності, яка обумовлена різноманітністю градацій фактора, а це саме і означає, що фактор діє набагато з більшою силою, залишаючи на частку випадкових впливів меншу частину загального різноманітності ознаки.

Таким чином, робимо висновок, що сила впливу фактора (факторів) в ДА (дисперсійному аналізі) вимірюється відношенням дисперсій часткових до загальної:

$$\eta_i^2 = \frac{C_i}{C_Y}$$

Через те що цей показник відображає головний закон розкладання загальних дисперсій і основну аддитивну властивість часткових дисперсій, а також складено з основних елементів дисперсійного аналізу, то ставлення однієї з факторіальних дисперсій (позначення C_V , C_A , C_B , C_{AB}) або випадкової дисперсії (позначено C_Z) до загальної (C_Y) можна назвати головним показником сили впливів факторів - організованих і неорганізованих.

Корінь квадратний з головного показника сили впливу в однофакторних комплексах це $\eta = \sqrt{\eta^2}$ - є пірсоновське кореляційне відношення, символ якого η перейшов на сучасний показник - сили впливу.

У 1-о -факторном комплексі визначаються два показника сили впливу: організованого фактора:

$$\eta_V^2 = \frac{C_V}{C_Y}$$

і неорганізованого фактора:

$$\eta_Z^2 = \frac{C_Z}{C_Y}$$

Сума цих показників дорівнює одиниці:

$$\eta_V^2 + \eta_Z^2 = 1$$

В 2-охфакторном комплексе определяются пять видов влияний:

Влияние первого фактора:

$$\eta_A^2 = \frac{C_A}{C_Y}$$

Влияние второго фактора:

$$\eta_B^2 = \frac{C_B}{C_Y}$$

Влияние сочетаний градаций обоих факторов:

$$\eta_{AB}^2 = \frac{C_{AB}}{C_Y}$$

Суммарное действие обоих факторов:

$$\eta_V^2 = \frac{C_V}{C_Y}$$

Действие случайных факторов:

$$\eta_Z^2 = \frac{C_Z}{C_Y}$$

Якщо ми спробуємо інтерпретувати показники четвертого і п'ятого впливів у двофакторний дисперсійний комплекс то будемо це проводити так само, як і в однофакторному комплексі: чим більше η_V^2 , следовательно, чим менше η_Z^2 , тим сильніше проявляється сумарна дія обох організованих факторів.

Інтерпретацію перших трьох впливів у двофакторному комплексі краще починати з показника який називається показник впливу поєднань градаций.

Цей показник позначаємо η_{AB}^2 завжди настільки більше нуля, наскільки сильно дія одного чинника залежить від дії (градаций) іншого.

Найменше значення це $\eta_{AB}^2 = 0$ ми можемо отримати якщо виходить, коли один фактор діє абсолютно однаково при будь-яких градаціях іншого.

Найбільше значення досліджуемого показника дорівнює показнику $\eta_{AB}^2 = \eta_V^2$ сумарного впливу організованих факторів: $\eta_{AB}^2 = \eta_V^2$. Це трапляється, коли дія 1-го фактора при 1-ій градації іншого чинника строго протилежна його дії при інших градаціях наступного фактора.

У таких крайніх випадках виходять дуже низькі показники часткових впливів першого фактора η_A^2 або другого η_B^2 , або того й іншого - вони наближаються до 0, але це не пов'язано зі слабкою дією кожного фактора окремо.

Якщо розглянути $\eta_A^2 \rightarrow 0, \eta_B^2 \rightarrow 0, \eta_{AB}^2 \rightarrow \eta_V^2$ то можна побачити як дія 1-го фактора настільки сильно залежить від дії другого, що стає неможливим вивчати і використовувати вплив першого фактора без урахування впливу 2-гого.

Показники сили впливу кожного фактора в 2-охфакторному комплексі які ми позначили η_A^2 та η_B^2 мають особливе значення, вони залежать від сили поєднання їх градацій: η_{AB}^2 .

Якщо показник поєднання градацій невеликий $\eta_{AB}^2 \rightarrow 0$, то показники часткових впливів факторів (η_A^2 та η_B^2) мають стандартне зручне значення: чим вони більші, тим сильніше вплив фактора.

Нужно тільки помнит, що сила кожного фактора окремо вимірюється в дисперсійному комплексі при усередненому дії градацій іншого чинника, що рівносильно известному вимозі вивчати варіанти впливів «при інших рівних умовах».

У тих випадках, коли зростає вплив поєднання градацій обох факторів ($\eta_{AB}^2 \rightarrow \eta_V^2$), вже не можливо за показниками (η_A^2 та η_B^2) судити повною мірою

про силу відповідних впливів. Як вже згадувалось, в таких випадках можливі дуже низькі показники сили статистичного впливу кожного фактора окремо при дуже помітному їх фізіологічному впливі на результативну ознаку.

В таких случаях сильна дія одного чинника має протилежний зміст у різних градаціях другого чинника. При усередненні таких протилежних дій виходить в більшій чи меншій степені нивелирование вимірювань сили впливу, а це у свою чергу призводить до зменшення показників сили приватного впливу кожного фактора окремо.

Точнісінька формула помилки основного показника сили впливу до сих пір не знайдена. В 1-факторних комплексах, коли помилка репрезентативності визначається тільки для 1-го показника факторіального впливу, допустимо користуватися таким варіантом загальної формули:

$$s_{\eta_v^2} = (1 - \eta_v^2) \cdot \frac{r - 1}{N - r}$$

У 2-факторних комплексах, якщо були розраховані дисперсії, можна використовувати формулу:

$$s_{\eta_i^2} = \eta_i^2 \cdot \frac{\sigma_z^2}{\sigma_i^2}, \text{ де використаний індекс } i - \text{це } V, A, B \text{ або } AB.$$

Якщо дисперсії не можливо розрахувати, то буде більш зручною загальна формула:

$$s_{\eta_i^2} = v_i \cdot \frac{\eta_z^2}{v_z}$$

В цьому випадку для 2-факторного комплексу проводиться ряд дій:

- 1) Знаходиться постійна величина $\frac{\eta_z^2}{v_z}$
- 2) Множиться на число ступенів свободи
- 3) По кожному впливу знаходяться помилки показника цих впливів для даного комплексу.

Якщо розглядати запропоновану помилку репрезентативності основного показника сили впливу то вона має істотні відмінності від звичайних помилок вибірових показників. Відношення основного показника сили впливу до цієї дії отримаємо його помилку:

$$F = \frac{\eta_v^2}{s_{\eta_v^2}}$$

Вона недорівнює критерію Стюдента (як звичайно), а дорівнює критерію Фішера при 2-ух ступенях свободи:

$$v1 = r - 1,$$

$$v2 = N - r.$$

Використання пропонованої помилки для визначення достовірності впливу дає точно такі ж результати, як і критерій Фішера.

Через те, що у запропонованої помилки можна визначити наближено довірчі інтервали основного показника сили впливу то це і є основна перевага, бо це не можна зробити за допомогою критерія Фішера.

Ці довірчі границі визначаються за формулою, в якій замість критерію Стюдента (t) введено критерій Фішера (F):

$$\eta_x^2 = \eta_x^2 \pm F_{st} \cdot s_{\eta_x^2}$$

Запропонована формула помилки володіє ще однією важливою властивістю: критерій достовірності, отриманий за допомогою цієї помилки, враховує відмінність в достовірності показників для комплексів різної структури, тобто - однакового обсягу, але з різним числом градацій (r) і з різною повторністю (n). Якщо, наприклад, було досліджено два комплекси однакового обсягу кількості $N = 100$ з рівним вибіровим показником сили впливу $\eta_x^2 = 0,5$, але з різною структурою $r1 = 1$, $n1 = 49$, $r2 = 49$, $n2 = 1$, то достовірність показника 1-ого комплексу повинна бути значно вище в порівнянні з вірогідністю показника 2-гого комплексу.

У 1-ому комплексі показник впливу отриманий при аналізі двох часткових середніх ($r_1 = 1$), з яких кожна підкріплена сорока дев'ятьма даними ($n_1 = 49$) і тому в набагато меншому ступені відображає випадковості у формуванні середніх величин.

У 2-ому комплексі, навпаки, показник впливу отриманий при аналізі п'ятидесятьох часткових середніх, з яких кожна усереднює лише дві ознаки і тому схильна до набагато більшим ступеням випадковості в залученні даних в градації.

Большое отличие в достоверности показника силы влияния в этих 2-ых комплексах в достатній мірі відображено в помилку репрезентативності:

$$s_{\eta_x^I}^I = (1 - 0,6) \cdot \frac{2-1}{100-2} = \frac{0,4}{98} = 0,004$$

$$s_{\eta_x^{II}}^{II} = (1 - 0,6) \cdot \frac{50-1}{100-50} = \frac{0,4 \cdot 49}{50} = 0,392$$

в критерии достоверности:

$$F_1 = \frac{\eta_x^2}{s_{\eta_x^I}^I} = \frac{0,6}{0,004} = 150 \quad ; \quad F_{st}^I = \{3,9 - 6,8 - 11,5\}$$

$$F_2 = \frac{\eta_x^2}{s_{\eta_x^{II}}^{II}} = \frac{0,6}{0,392} = 1,5 \quad ; \quad F_{st}^{II} = \{1,6 - 2,0 - 2,7\}$$

и в доверительных границах:

$$\Delta_1 = s_{\eta_x^I}^I \cdot F_{st}^I = 0,004 \cdot 3,9 = 0,016 \quad ;$$

$$\eta_x^2 = \eta_x^2 \pm \Delta_1 = 0,6 \pm 0,016 = 0,584 \div 0,616$$

$$\Delta_2 = s_{\eta_x^{II}}^{II} \cdot F_{st}^{II} = 0,392 \cdot 1,6 = 0,627 \quad ;$$

$$\eta_x^2 = \eta_x^2 \pm \Delta_2 = 0,6 \pm 0,627 = -0,027 \div 1,227$$

Тому слід зазначити, що більша відмінність комплексів по достовірності їх показників абсолютно не учитивається звичайною помилкою кореляційних відносини. Для обох щойно розібраних комплексів помилка репрезентативності кореляційного відносини буде однаковою:

$$s_{\eta} = \sqrt{\frac{1 - \eta^2}{N - 2}} = \sqrt{\frac{1 - 0,6^2}{98}} = 0,081$$

Основним показником сили впливу буде дорівнювати частці одного доданка від всієї суми доданків. Цей показник дорівнює квадрату кореляційних відносини. Тобто ми маємо дві причини чому показник сили впливу завжди більше нуля, він не може бути негативним (менше нуля). Найменша його величина $\eta_x^2 = 0$, коли всі часткові середні по градаціях комплексу виявилися однаковими, рівними загальної середньої. Найбільша величина показника $\eta_x^2 = 1$, коли всі дані всередині кожної градації однакові і рівні своїй приватної середньої.

Тільки в одному випадку основний показник сили впливу може ставати менше нуля або більше одиниці, це випадок коли при визначенні довірчих меж генерального параметра на основі нечисленного вибіркового комплексу, при великій різноманітності значень досліджуваного ознаки.

У всіх інших випадках (коли не визначаються довірчі кордону) отримання показника сили впливу негативного або більше одиниці завжди вказує або на помилку рахунку, або на порочне метод визначення сили впливу.

Основний показник сили впливу, отриманий у вибіркового дослідженні, характеризує, перш за все, ту ступінь впливу, яка реально, насправді, проявилася в групі досліджених об'єктів, і як первинний факт підлягає безпосередньому вивченню і включенню в загальну ланцюг спостережень, зіставлень і розкриття причин.

У той же час матеріали вибіркового комплексу, в якому визначено основний показник сили впливу, можуть бути використані також і для оцінки відповідного генерального параметра, т. Е. Ступеня впливу, властивої загальному комплексу генеральних сукупностей, відповідних градаціях вибіркового комплексу.

Оцінка генерального параметра не може бути проведена шляхом простого прирівнювання його до того показника сили впливу, який виявлений у вибіркового комплексі. Прогноз генеральних параметрів сили впливів за

вибірковими показниками завжди може бути зроблений з більшою або меншою похибкою, неминуchoю при аналізі будь-якого вибіркового комплексу.

Вийшло в комплексі різноманітність приватних середніх ніколи точно не відповідає різноманітності генеральних середніх внаслідок звичайних помилок репрезентативності при випадковому наборі об'єктів і даних в градації.

Ця неточність в крайніх випадках може привести до великої різноманітності вибірових приватних середніх при дуже незначних відмінностях або навіть повної рівності відповідних генеральних середніх по градаціях комплексу. У подібних випадках вибіровий показник сили впливу дає перебільшену характеристику сили впливу в генеральному комплексі.

Можлива й інша крайня похибка, коли випадковості набору об'єктів і даних в градації вибірового комплексу призведуть до дуже малої різноманітності вибірових часткових середніх при великій різноманітності відповідних генеральних середніх. У подібних випадках вибіровий показник сили впливів дасть зменшений прогноз генерального параметра сили впливу.

Ошибки в оцінці генерального параметра з вибірового показника ε у кожному вибіровому дослідженні, в тому числі і вибірового дисперсійному аналізу. Тому, як і у всякому вибіровому дослідженні, при дисперсійному аналізі сили впливів визначаються показники, що допомагають з'ясувати можливу величину помилок прогнозу генеральних параметрів за вибіровими показниками.

Перелік помилок репрезентативності в ДА проводиться в формі критерію достовірності вибіровому показнику і довірчих меж генерального параметра сили впливу. В основі врахування цих помилок репрезентативності лежать наступні закономірності.

Отличие різноманітності вибірових середніх від різноманітності відповідних генеральних середніх не може бути безмежним. Наприклад, за однакової кількості генеральних середніх різноманітність відповідних вибірових середніх не може бути більше певної величини, яку можна встановити при проведенні аналізу вибірових дисперсійних комплексів.

При повній рівності генеральних часткових середніх різноманітність вибірових часткових середніх не може бути більше особливого показника - критерія Фішера при заданій ймовірності безпомилкових прогнозів.

Якщо різноманітність часткових середніх в вибіровому комплексі не досягає критерія Фішера, це означає що це вибірова різноманітність могла вийти в порядку випадкових відхилень від нульової різноманітності відповідних генеральних середніх. У таких випадках вибіркою показник сили впливів недостовірний, а прогноз генерального параметра невизначений, тому що не відкидає і не підтверджує впливу фактора в генеральному комплексі, при масовому застосуванні чинника. В таких случаях, при недостовірності показника сили впливу, емпіричний показник цілком прийнятий при характеристиці впливу тільки в межах вивченого комплексу і не може бути використаний для встановлення наявності або відсутності впливу в генеральному комплексі.

Якщо перелік часткових середніх в вибіровому комплексі дорівнює чи перевищує критерій Фішера, значить, це вибірова різноманітність, а вона вже не могла вийти внаслідок випадкових відхилень від різноманітності відповідних генеральних середніх. З цієї причини різноманітність часткових середніх перейшло допустимий поріг, який визначається критерієм Фішера, що і вказало на достовірність досліджуваного впливу.

Сопоставление емпіричного критерію з його стандартними значеннями може дати два різних результати.

Емпіричний критерій не досягає свого звичайного значення, взятого відповідно до встановленого порогом ймовірності безпомилкових прогнозів.

У таких випадках для необхідної ймовірності неможливо зробити висновки як про рівність, так і про відмінність відповідних генеральних середніх, так як маленька різноманітність вибірових часткових середніх може вийти при будь-якому (великому чи малому, або нульовому) розмаїтті генеральних середніх по градаціях комплексу. А це у свою чергу означає, що в таких випадках не можна дати певного прогнозу про вплив генерального фактора: залишається

нез'ясованим, можна чи ні очікувати із установленою ймовірністю, що при масовому застосуванні чинника виходять результати, подібні до тих, які отримані в вибірковому комплексі, звичайно, при вивчених градаціях фактора і за даних умов.

Потрібно запобігати двом помилковим думкам про недостовірність показника сили впливу. Неправильною буде думка, що отримання недостовірного показника сили впливу вказує на те, що «впливу взагалі немає», це означає що вплив відсутній в генеральних сукупностях.

Отримання недостовірного показника не може підтверджувати, чи заперечувати генеральний вплив.

Не можна, також вважати, що при отриманні недостовірного показника сили впливу в проведеному дослідженні взагалі нічого не отримано і це дослідження проведено без будь-якої користі. Це велика помилка. Та міра впливу, яка при цьому отримана, цілком відноситься до групи вивчених об'єктів і як експериментальний факт повинна бути врахована і в даному, і в інших, і в подальших роботах.

У деяких випадках вивчення сили впливу проводиться тільки для певної обмеженої групи об'єктів, з яких і складається дисперсійний комплекс. У таких випадках не ставиться завдання визначити силу генерального впливу, і емпіричний показник сили впливу набуває повне значення без визначення його достовірності.

У деяких дослідженнях саме недостовірність показника сили впливу, визначена за прямим відношенню дисперсій, дає відповідь на основне питання цього дослідження. Так буває в тих випадках, коли недостовірність за прямим відношенню дисперсій не спростовує подібності досліджуваних особин за їх особистими якостями або спадковим здібностям як представників однієї лінії.

Визначеність прогнозів набуває чинності достовірності, якщо при недостовірності за прямим відношенню дисперсій малий вплив (а значить, велику схожість градацій) виявляється достовірним по зворотному відношенню

$$F = \frac{\sigma_z^2}{\sigma_v^2} |$$

Емпіричний критерій дорівнює або перевищує своє стандартне значення з необхідною ймовірністю.

У таких випадках можливий певний прогноз: генеральні середні по градаціях комплексу неоднакові і їх різноманітність подібно до того, яке спостерігалось в вибіркового комплексу. Різноманітність приватних середніх в вибіркового комплексу тепер уже не може бути пояснено тільки випадковостями вибіркового дослідження.

Достовірний вплив означає, що вивчений фактор при його масовому застосуванні в певних градаціях і в умовах буде впливати на результативний ознака з ймовірністю, знайденої при оцінки достовірності його сили впливу.

Більш детально можна прочитати у статті [3].

2.2 Методи коваріційного аналізу

Взагалі для аналізу даних можуть застосовуватися різні методи в залежності від поставлених задач та цілей. Статистичні методи аналізу даних призначені для їх спрощення, виявлення взаємозв'язків і структур.

Розглянемо взагалішому що таке статистичні методи:

Статистичні методи – це методи аналізу статистичних даних, тобто яких назбирали певну велику кількість. За своїм походженням вони поділяються на кількісні і категоріальні.

Кількісні (метричні) дані є безперервними даними за своєю структурою. Ці дані можна поміряти або виміряти за допомогою інтервальної шкали (числова шкала, кількісно рівні проміжки якої відображають рівні проміжки між значеннями вимірюваних характеристик), або за допомогою шкали відносин (крім відстані визначений і порядок значень).

Категоріальні (Не метричні) дані - це якісні дані тобто у відповідність складно поставити числа, вони приймають значення з обмеженим числом унікальних значень і категорій. Існує два види категоріальних даних: номінальні - використовується для нумерації об'єктів і порядкові - дані, для яких існує природний порядок категорій.

Статистичні методи поділяються на 1-о- і багатовимірні методи. Одновимірні методи використовуються тоді, коли всі елементи вибірки оцінюються єдиним вимірником або якщо цих вимірників кілька для кожного елемента, але кожна змінна аналізується при цьому окремо від усіх інших.

Univariate techniques – це одновимірні статистичні методи статистичного аналізу даних у випадках, якщо існує єдиний вимірювач для оцінки кожного елемента вибірки або якщо ці лінійок декілька, але кожна змінна аналізується окремо від усіх інших.

Одновимірні методи (рис. 2.1) можна класифікувати на базі того, які дані аналізуються: метричні або не метричні. Metric data вимірюються по інтервальній або відносній шкалі. Nonmetric data оцінюються за номінальною або порядковою шкалою.

Потім ці методи ділять на класи на основі того, скільки вибірок - одна, дві або більше - аналізується в ході дослідження. Варто відзначити, що число вибірок визначається тим, як ведеться робота з даними для конкретного аналізу, а не тим, яким способом збиралися дані.



Рисунок 2.1 Класифікація одновимірних статистичних методів

Порівняння середніх це один із способів виявлення взаємозв'язку між всіма змінними - ознаки, що характеризують досліджувану сукупність об'єктів. Якщо при розбитті об'єктів дослідження на підгрупи за допомогою категоріальної незалежної змінної (категоріального предиктор) підтверджена гіпотеза про не рівність середніх деякої залежної змінної (так званий відгук) в підгрупах, то це означає, що існує стохастична взаємозв'язок між залежною змінною і категоріальним предиктором.

Так, наприклад, якщо встановлено, що не вірна гіпотеза про рівність середніх показників фізичного і інтелектуального розвитку дітей в групах матерів, курили і не палили під час вагітності, то це означає, що існує

залежність між курінням матері дитини в період вагітності і його інтелектуальним і фізичним розвитком. Найбільш загальний метод порівняння середніх - дисперсійний аналіз (ANOVA). ДА який більш детально розбирається вище можна визначити як параметричний, статистичний метод, призначений для оцінки впливу різних категоріальних факторів (предикторів) на результат експерименту, а також для подальшого планування експериментів. Тому в дисперсионном аналізі можна досліджувати залежність кількісної ознаки – відгуку від одного або декількох факторів і їх комбінацій. У термінології пакета STATISTICA категоріальні предиктори і їх комбінації називаються ефектами. Дисперсійний аналіз дозволяє побудувати регресійну модель залежності відгуку від ефектів

Регресійний аналіз дозволяє досліджувати залежність кількісної ознаки /відгуку від однієї, або декількох незалежних кількісних факторів (предикторів) і побудувати математичну модель залежності, яка називається рівнянням регресії.

Якщо порівняти, то дисперсійний та регресійний аналіз коваріаційний аналіз (ANCOVA) - розділ аналізу даних, мета якого дослідити характер взаємозв'язку між залежною величиною - відгуком і набором кількісних і якісних незалежних величин – предикторів і побудувати регресійну модель, тобто він є синтезом регресійного і дисперсійного аналізу. Не залежні кількісні змінні, що відносяться до інтервальної шкали або до шкали відносин (метричної), називаються коваріатами. Тому, як коваріат має використовуватися безперервна величина, або дискретна (порядкова) з великою кількістю значень.

Якщо в дисперсійному аналізі оцінюється ступінь випадкової мінливості відгуку з боку ефектів - категоріальних предикторів і їх комбінацій, то в коваріаційного аналізу оцінюється ступінь мінливості відгуку також і з боку безперервних предикторів, званих коваріатами.

Відносно коваріат робляться припущення про те, що вони поряд з ефектами обумовлюють деяку частку варіації (мінливості) залежною змінною. Якщо

ступінь мінливості відгуку від коваріат велика, то ми говоримо про статистично значущому впливі коваріат на відгук.

Коваріаційний аналіз на відміну від загальні регресивні моделі дозволяє оцінити статистичну значущість впливу коваріат на відгук при виключенні мінливості відгуку з боку ефектів, тобто в припущенні, що ефекти приймають фіксовані значення; і при виключенні мінливості відгуку з боку коваріати, тобто в припущенні, що коваріата приймає фіксоване значення.

В силу викладених обставин, коваріаційний аналіз є сенс проводити, якщо попередньо встановлено наявність статистично значущої взаємозв'язку між відгуком і ефектами - категоріальним предикторами і їх комбінацією за допомогою непараметричної кореляції або дисперсійним аналізом. Якщо значущою взаємозв'язку немає, то для оцінки ступеня впливу коваріат на відгук досить скористатися звичайним регресійний аналізом.

Аналогічно, за допомогою регресійного або кореляційного аналізу доцільно встановити наявність статистичного взаємозв'язку між відгуком і коваріатами. Якщо такого взаємозв'язку немає, то для оцінки впливу ефектів на відгук досить скористатися звичайним дисперсійним аналізом.

Коваріаційний аналіз - сукупність методів математичної статистики, що відносяться до аналізу моделей залежності середнього значення деякої випадкової величини Y одночасно від набору (основних) якісних факторів F^i (супутніх) кількісних факторів X . Фактори F задають поєднання умов, при яких були отримані спостереження X, Y , і описуються за допомогою індикаторних змінних, причому серед супутствующих і індикаторних змінних можуть бути як випадкові, так і не випадкові (контрольовані в експерименті).

Якщо випадкова величина Y є вектором, то говорять про багатовимірному коваріаційного аналізу.

Коваріаційний аналіз часто застосовують перед дисперсійним аналізом, щоб перевірити гомогенність (однорідність, представництво) вибірки спостережень X, Y за всіма супутнім факторам.

Більшість теоретичних та прикладних проблем коваріаційного аналізу відносяться до лінійних моделей. Зокрема, якщо аналізуються n спостережень Y_1, \dots, Y_n з p супутніми змінними $(X = (x^{(1)}, \dots, x^{(p)}))$, k можливими типами умов експерименту $(F = (f_1, \dots, f_k))$, то лінійна модель відповідного коваріаційного аналізу задається рівнянням:

$$Y_i = \sum_{j=1}^k f_{ij} \theta_j + \sum_{j=1}^p \beta_j x_i^{(j)} + \epsilon_{ij}$$

Де зазначені $i = 1, \dots, n$, індикаторні змінні, а f_{ij} рівні одиниці, якщо j -е умова експерименту мала місце при спостереженні Y_i , і рівні нулю в іншому випадку. Коефіцієнти θ_j визначають ефект впливу j -ї умови, $x_i^{(j)}$ - значення супутньої змінної $x^{(j)}$, при якій отримано спостереження Y_i , β_j - значення відповідних коефіцієнтів регресії Y за $x^{(j)}$, ϵ_{ij} - незалежні випадкові помилки з нульовим математичним очікуванням.

Вищенаведена формула задає лінійну модель 1-офакторного коваріаційного аналізу з p незалежними змінними і k рівнями фактора. Якщо включити в модель додаткових чинників то в правій частині рівняння з'являться доданки, що відповідають за ефекти рівнів нововведених в модель факторів. Коефіцієнти регресії в наведеній формулі не залежать від якісних факторів. Це включає припущення, що лінійна залежність має однакові коефіцієнти для кожного значення якісного фактора.

Основне назначение коваріаційного аналізу - використання в побудові статистичних оцінок $\theta_1, \dots, \theta_k; \beta_1, \dots, \beta_p$ і статистичних критеріїв для перевірки різних гіпотез щодо значень цих параметрів. Якщо в моделі були такі $\beta_1 = \dots = \beta_p = 0$, то вийде модель дисперсійного аналізу, якщо ж виключити вплив не кількісних факторів (покласти $\theta_1 = \dots = \theta_k = 0$), то вийде модель регресійного аналізу.

Основною гіпотезою, що перевіряється в коваріаційного аналізу, є

$$H_0: \theta_1 = \theta_2 = \dots = \theta_k.$$

Якщо йдеться про одну змінну ($p=1$) то цю гіпотезу можна інтерпретувати в такий спосіб:

За припущеннями лінійної моделі коваріаційного аналізу для кожного рівня фактора криві регресії залежної змінної Y на супутствующую змінну x паралельні. Гіпотеза H_0 передбачає, що ці криві збігаються.

Зазвичай ця гіпотеза перевіряється за допомогою критерію Фішера в результаті відомості поставленого завдання до завдань дисперсійного аналізу

2.2 Лінійна імовірнісна модель

ЛІМ – це лінійна імовірнісна модель формою якої є лінійна регресія, в якій залежна змінна набуває значення 0 або 1, якщо подія відбулась чи не відбулась.

Запишемо модель у формальному математичному вигляді:

$$y(k) = b_0 + b_1 x_1(k) + b_2 x_2(k) + \dots + b_m x_m(k) + \varepsilon(k), \quad (2.1)$$

де y – залежна змінна, значення якої відповідає результату повернення кредиту;

b_0, \dots, b_m – коефіцієнти (параметри) рівняння регресії, які оцінюються за даними, які характеризують клієнтів;

x_1, \dots, x_m – пояснювальні змінні (характеристики клієнта);

$\varepsilon(k)$ – випадковий процес, зумовлений наявністю не вимірюваних збурень, а також помилок оцінювання структури і параметрів моделі;

k – ідентифікатор клієнта.

Якщо ми застосовуємо лінійну імовірнісну модель, то треба бути обережними бо є ряд недоліків:

а) існують випадки коли залежна змінна може набувати значення поза межами інтервалу від $[0, 1]$;

б) ЛІМ можна застосовувати до змінних, які приймають тільки неперервні значення;

в) будь-які процеси – кредитування, чи виробництва частіше характеризуються нелінійними залежностями, а це потребує застосування вже моделей зовсім інших структур.

Тому існують можливі способи виправлення вищезазначених проблем:

а) За допомогою обраного порогу відсікання (наприклад, 0.5) можна трансформувати результуючі значення залежної змінної:

$$\begin{cases} y = 0, \text{ якщо } y < 0.5; \\ y = 1, \text{ якщо } y \geq 0.5. \end{cases}$$

б) Кодування категоріальних змінних за допомогою:

- 1) порядкової нумерації категорій;
- 2) значень коефіцієнту зваженої сукупності (WOE);

в) Використання замість ЛПМ логіт- та пробіт-моделі.

2.4 Збір даних

Самий перший з основних етапів побудови будь-яких моделей є збір достатньої кількості репрезентативної вибірки даних. В цілому точність прогнозу та успіх розробленої системи залежить від якості вихідних даних. Для побудови прогнозуючих та оцінюючих моделей необхідно використовувати надійні і очищені дані з мінімально допустимою кількістю «поганих» і «хороших» записів. Об'єм необхідних даних визначається за допомогою вимог статистичної значущості і випадковості, але взагалі може бути різним.

Для вирішення практичних задач розробки статистичних моделей, експерти рекомендують використовувати не менше 2000 «поганих» та 2000 «хороших» записів про події, які вибираються випадковим чином з загальної історії роботи лінії. Потрібно убрати з вихідних даних інформацію про певний тип змін. Це можуть бути нетипові зміни, такі як техзапуски чи створення нового продукту тощо. Також треба виключити з бази зміни з дуже маленькою виробітком, тому що це ознака скоріше більш значної поламки на яку ми вплинути не можемо.

2.5 Визначення та обробка пропусків

Зазвичай історичні дані характеризуються відсутністю деяких необхідних значень або, навпаки, присутністю значень, що є некоректними та не можуть описувати ту чи іншу характеристику. Це можуть бути поля, значення яких більше не використовуються, або не були зафіксовані, або ж які були недоступні чи не були заповнені позичальниками тобто пропущенні значення; а також неправильно введені дані, викиди або значення, що дуже виділяються, тобто помилкові, некоректні дані.

Є кілька методів для обробки даних з такими значеннями, наприклад:

а) виключити з аналізу всі дані з пропущеними значеннями, оскільки аналіз ведеться по всіх змінних. У випадку роботи з реальними фінансовими даними такий спосіб в більшості випадків вилучає занадто багато даних;

б) виключити з моделі характеристики чи записи, в яких доля пропущених значень є більше заданого порога (наприклад, 20%);

в) включити до аналізу нову характеристику (ідентифікатор), що відображає наявність пропуску по атрибуту клієнта;

г) замінити пропущені значення, базуючись на середньому значенні, або прогнозуванні (наприклад дерева рішень чи регресійні методи), або статистичних спеціалізованих методів (синтетичний розподіл)

2.6 Оцінка кваліфікації співробітника

Для того щоб знаходити залежності між рівнем ЗВ, чи тривалостями простоїв від співробітників потрібно відокремитись від працюючих конкретних людей, а якось кластеризувати співробітників.

Які дані в нас є:

- Пол (м/ж)

- Вік (кількість повних років)
- Стаж роботи (кількість повних років у компанії)
- Освіта (неповна вища/повна вища/середня/професійно-технічна)
- Бали за іспит

Іспит поділен на 5 частин:

- Теорітичні питання. Основи технології кондитерського виробництва
- Теорітичні питання. Експлуатація, ремонт, техобслуговування
- Теорітичні питання. Загальні питання у сфері
- Практичні завдання. Завдання на збірку розбірку деталей
- Практичні завдання. Ескізування

За результатами проходження іспитів ми можемо кластеризувати наших співробітників. Це дає змогу на пошук залежностей між так званою кваліфікацією та витратами.

Наведемо декілька цікавих та корисних фактів. На рис. 2.6.1 наведено дані щодо розподілу кількості працівників в залежності від віку. Бачимо преобладання робітників віком від 30 до 35 років.

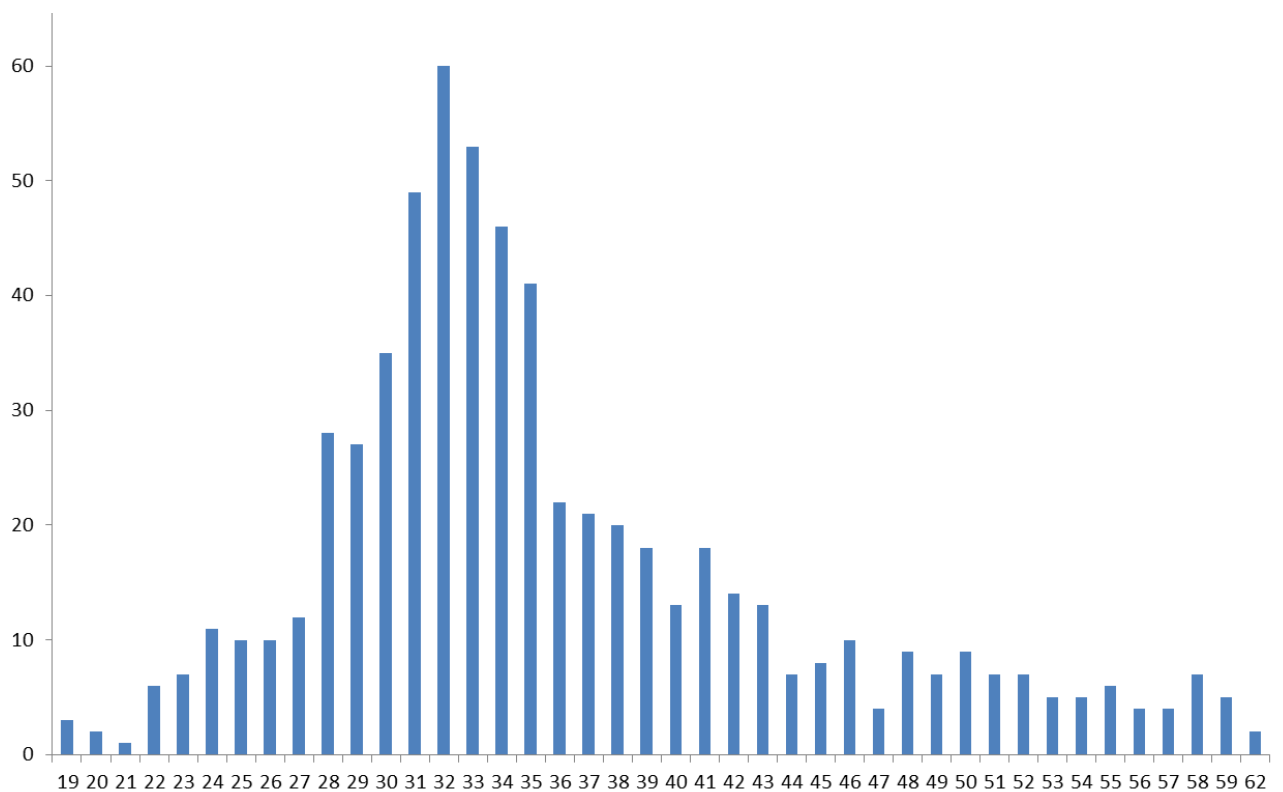


Рисунок 2.6.1 Розподіл кількості робітників за віком

На рисунку 2.6.2 залежність кількості загально набраних балів від віку працівника. Бачимо також трохи зміщення до 30-35 років. Цікаво побудувати залежність між теоритичними балами та практичними і віком працівника. Першою є гіпотеза про високі теоритичні бали у молодих людей, та високі практичні бали у старших людей. Це продемонстровано на рис 2.6.3

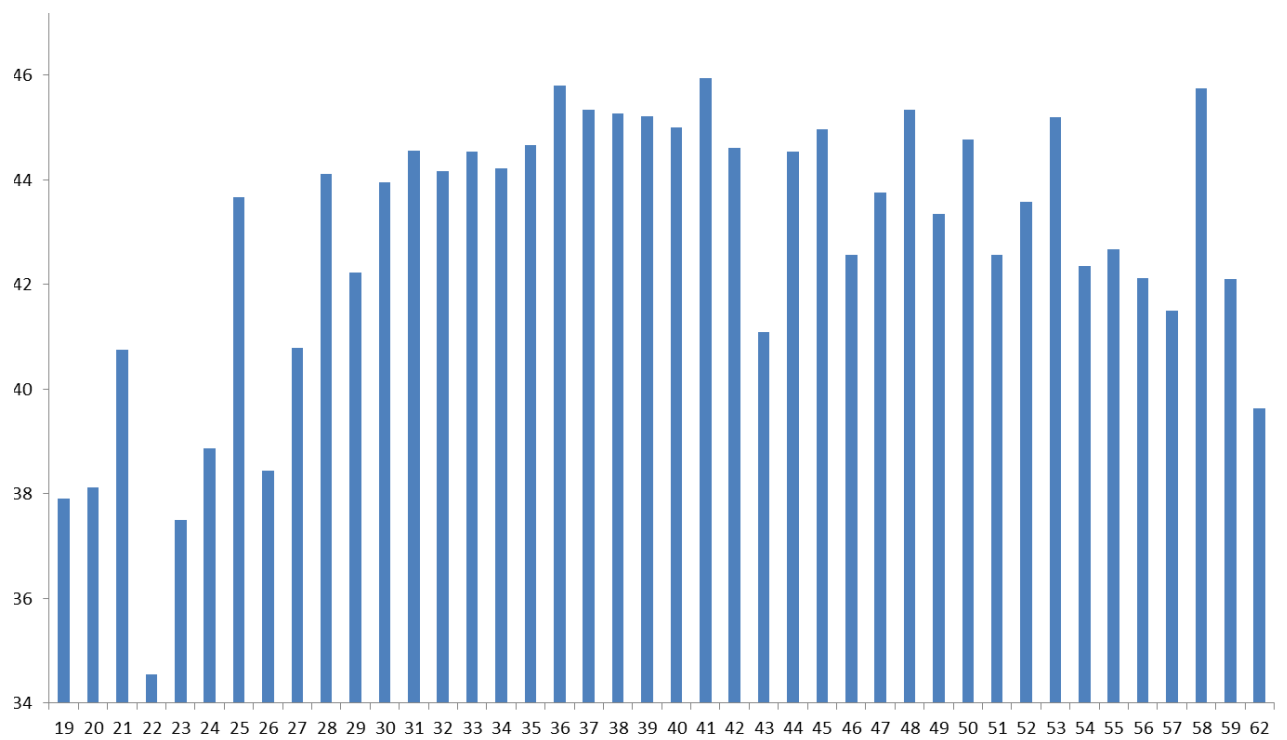


Рисунок 2.6.2 Залежність кількості загально набраних балів від віку працівника

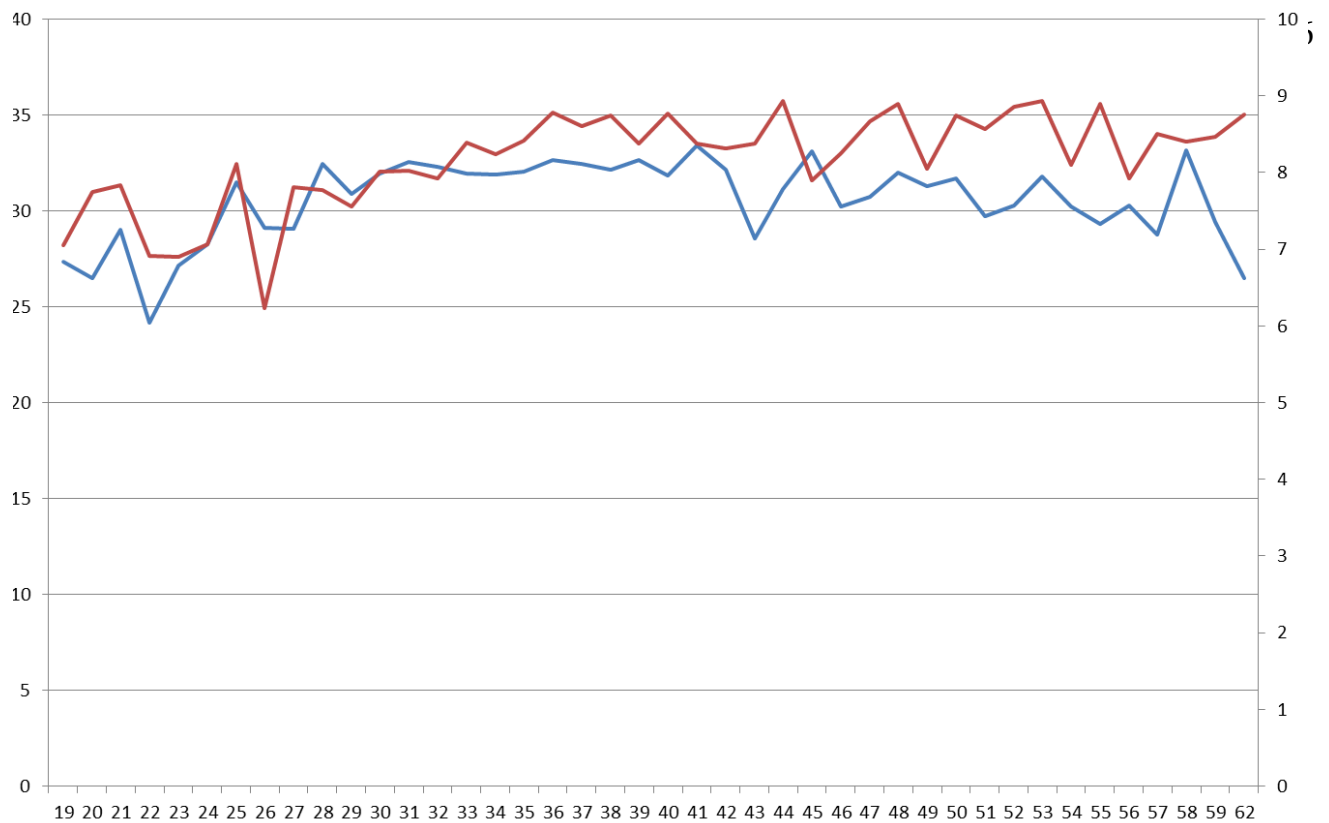


Рисунок 2.6.2 Залежність кількості теоритичних та практичних набраних балів від віку працівника. Червоне – практичні бали, Синій - теоритичні

Як бачимо, є тенденція більш досвідчених у практиці більш старших людей, а от гіпотези про високі теоритичні бали у молодих не справдилася. Це може бути спричинено різними факторами. Наприклад, більш старші люди мають більше зобов'язань, тому відповідальніше відносяться до тестувань та іспитів.

Наступний розподіл який подано на рисунку 2.6.3 є також дуже цікавим. Це взаємозв'язок стажу та оцінок. Бачимо, що чим більший стаж – тим більше показники. Випадок з 13 роками стажу та низькими показниками оцінок є виключним випадком. Статистичною неточністю.

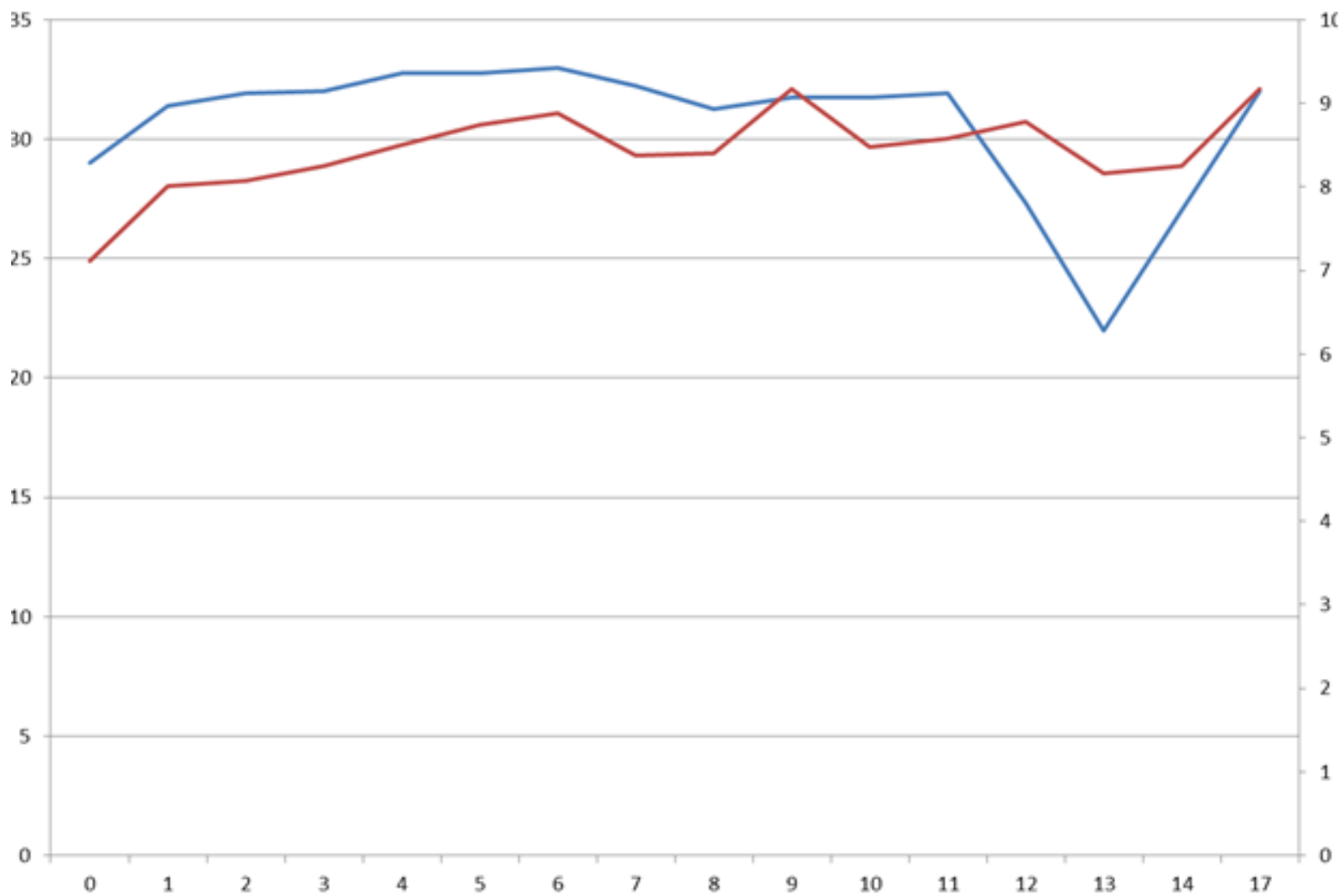


Рисунок 2.6.3 Залежність кількості теоритичних та практичних набраних балів від стажу працівника. Червоне – практичні бали, Синій - теоритичні

Рисунок 2.6.4 показує взагаліному які похибки при оцінюванні в нас можуть бути. От наприклад, ми бачимо, що із стажем роботи в 13 років в нас одна людина. Тобто ця яма на попередніх графіках нам не дуже інформативна. З цього ми робимо висновок, що доцільно сгрупувати людей на кластери стажів.

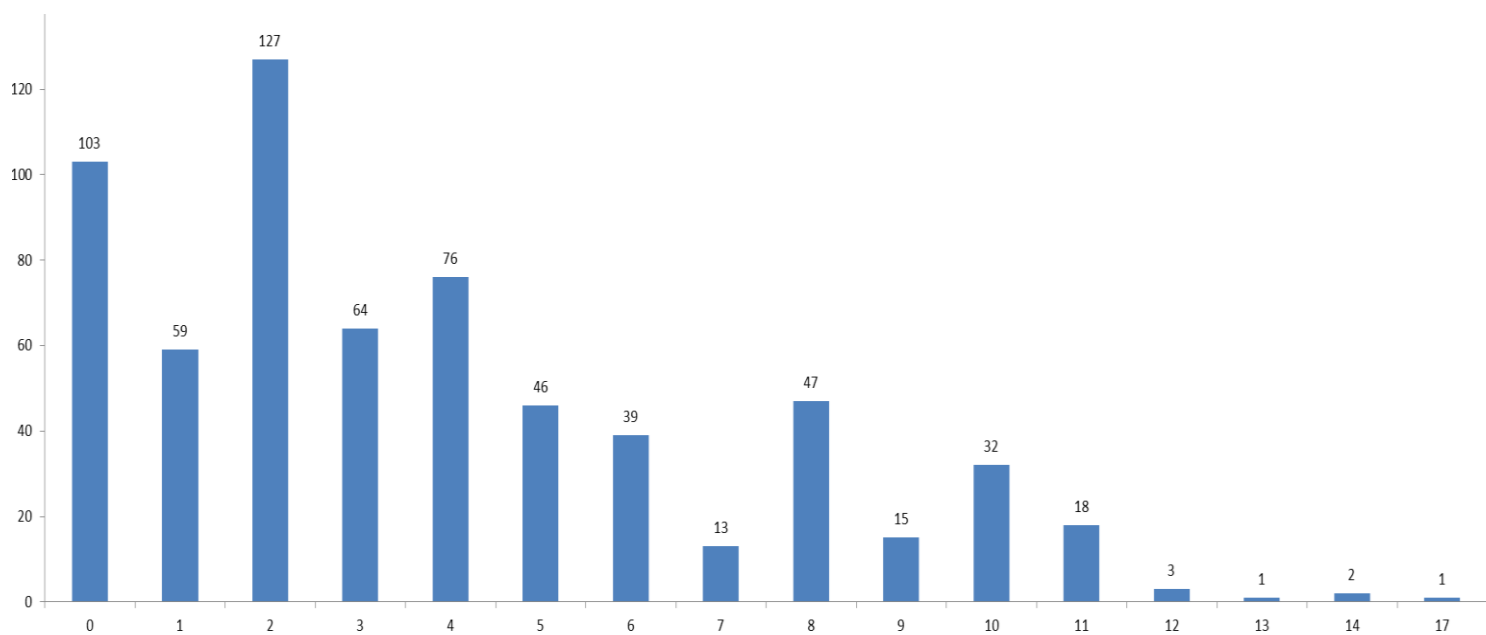


Рисунок 2.6.4 Кількість працюючих людей та відповідний стаж роботи.

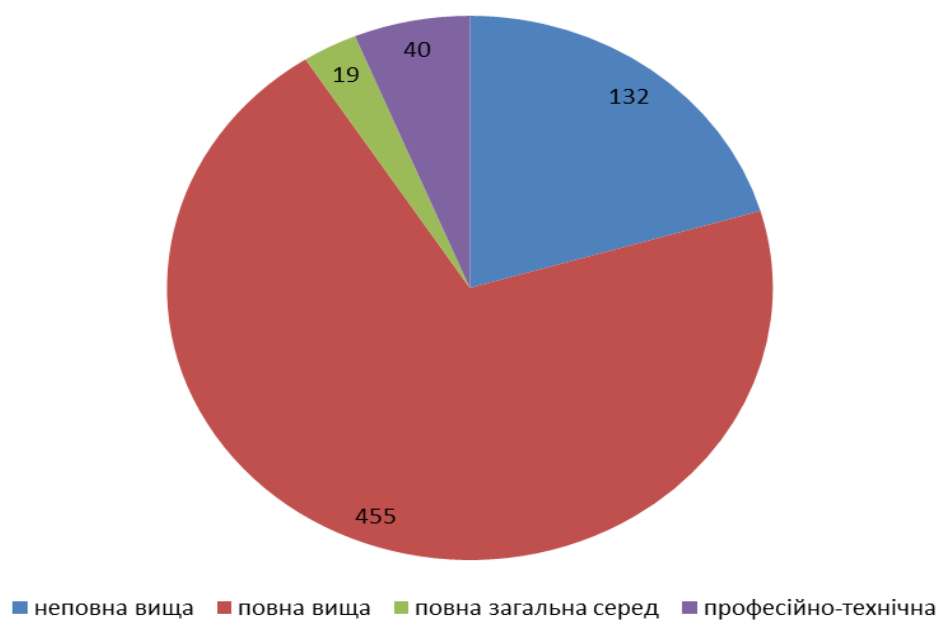


Рисунок 2.6.5 Кількість працюючих людей за відповідними рівнями освіти.

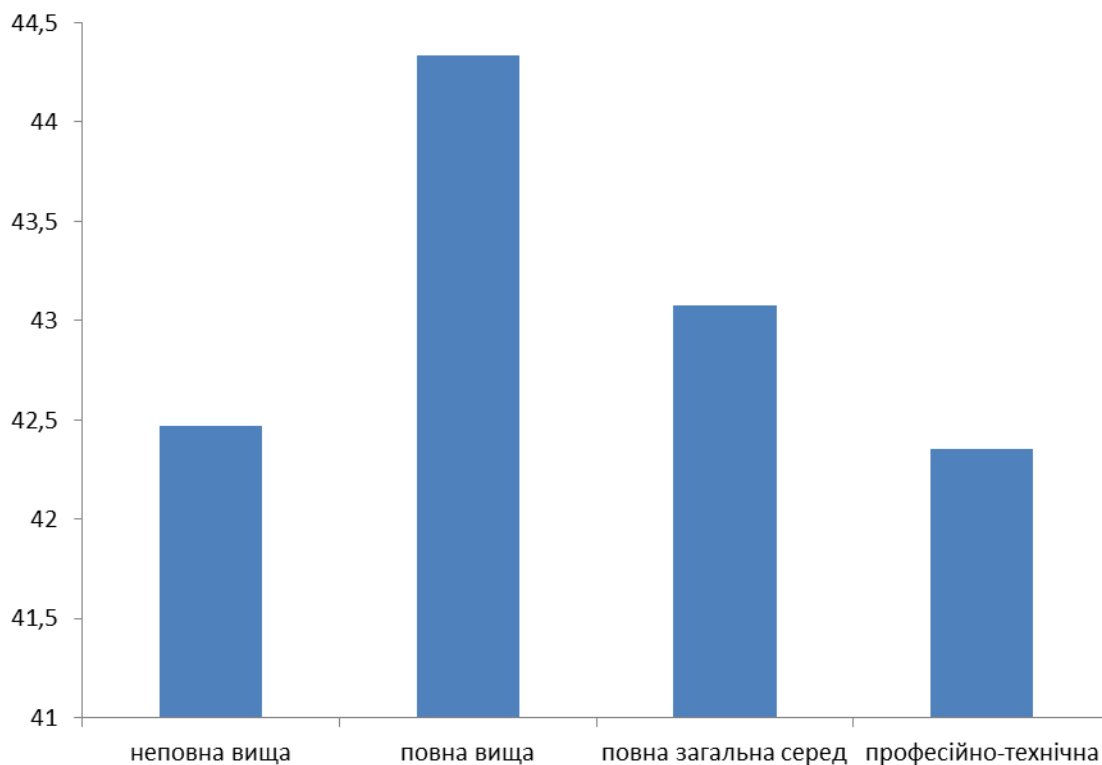


Рисунок 2.6.6 Середній бал групи за відповідними рівнями освіти

Рисунок 2.6.6 є наглядним доказом, що повна вища освіта є деяким аргументом у житті. Як би не навчався учень, всеодно щось відкладається у голові.

А от на рисунку 2.6.7 гарно бачити розподіли яки вище були зазначені але погано просматриварились через малу кількість робітників певного віку.

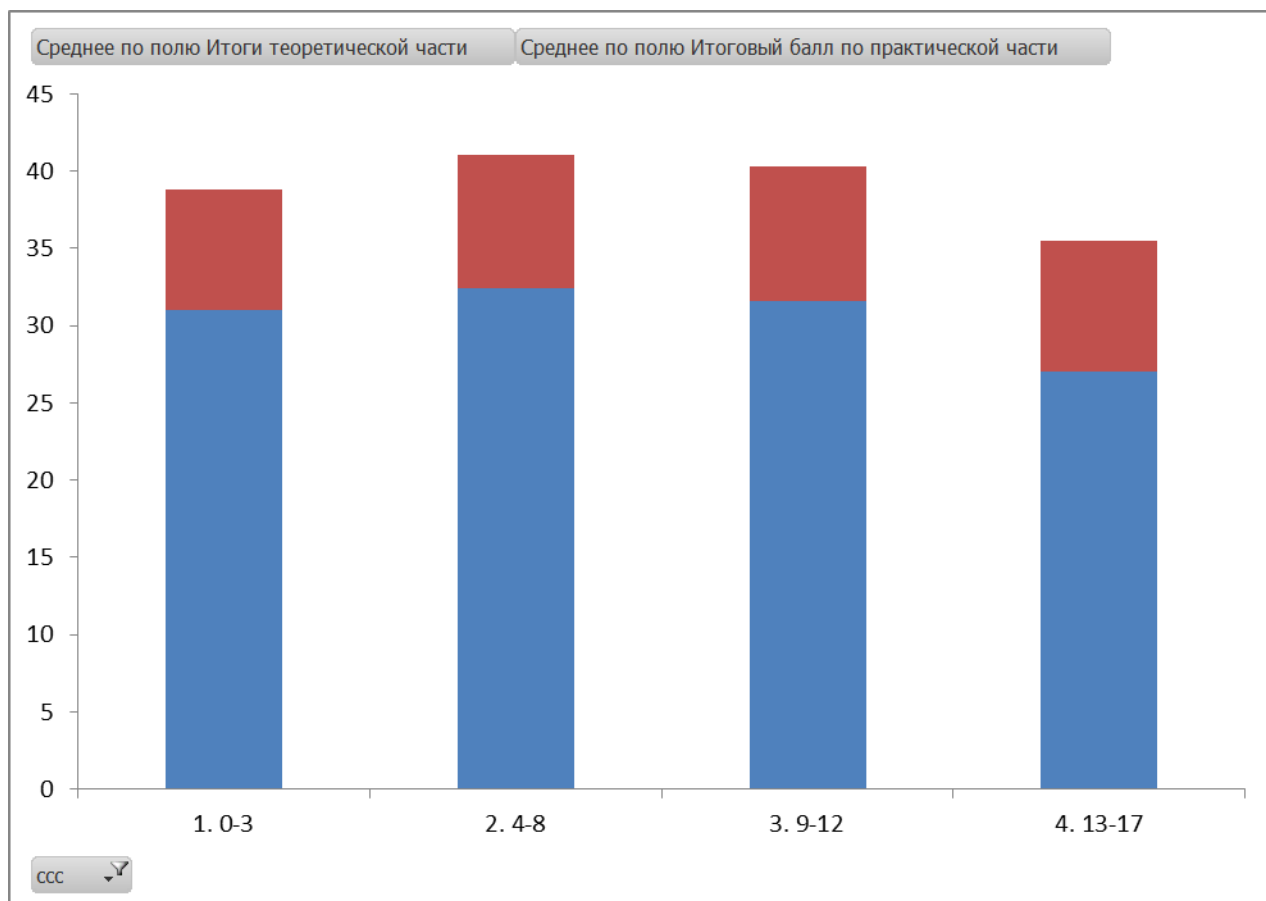


Рисунок 2.6.7 Середні бали по практичній та теоритичній частині в залежності від кластеру стажу роботи

Висновки до розділу 2

В другому розділі ми розглянули існуючі математичні методи прогнозування, які у подальшому ми будемо використовувати для оцінювання факторів виробництва, а саме дисперсійний аналіз, лінійна імовірнісна модель, коваріаційний аналіз, байєсовські мережі. Метод дерева рішень є досить зручним способом візуалізації процесу прийняття рішень. За допомогою нього ми бачимо по яких гілках ми йдемо тобто на якому етапі зараз ми знаходимося.

Для можелей у яких присутні великі об'єми даних, які можуть приймати широкий діапазон значень, побудована модель дерева рішень буде досить складною. Для цього було розглянуто різні алгоритми побудови дерев рішень. Було розглянуто ЛІМ (лінійну імовірнісну модель) у формі лінійної регресії, вона є досить простою у побудові та розрахунках. Дана модель має ряд недоліків. Було запропоновано способи вирішення певних недоліків. У розділі наведено алгоритми, які використовують для побудови та навчання логістичної регресії.

Досліджено процес попереднього аналізу і обробки даних для того щоб правильно проводити аналіз даних, який включає в себе: збір даних, обробку пропусків, визначення цільової змінної, відбір найбільш значущих змінних. Для використання різних математичних методів запропоновано метод перетворення категоріальних змінних у числові, наприклад для побудови регресійних моделей. Також були розглянуті критерії оцінки якості отриманих прогнозуючих моделей: загальна точність моделі, помилки 1-го й 2-го роду. Останні особливо важливі з погляду фінансової установи, оскільки помилка 1-го роду означає, що підприємство втратить певну суму грошей, а помилка 2-го роду означає, що підприємство не доодержить деякий прибуток.

РОЗДІЛ 3 СИСТЕМА ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ ВИЗНАЧЕННЯ ПРИЧИН ВИТРАТ НА ВИРОБНИЧІЙ ЛІНІЇ

3.1 Перевірка на нормальність розподілу зворотніх відходів

Для того щоб почати дисперсійний аналіз ми повинні бути впевнені, що зворотні відходи піддаються гаусівському розподілу. Для цього побудуємо розподіл зворотніх відходів рис 3.1. Основні показники на рис 3.2

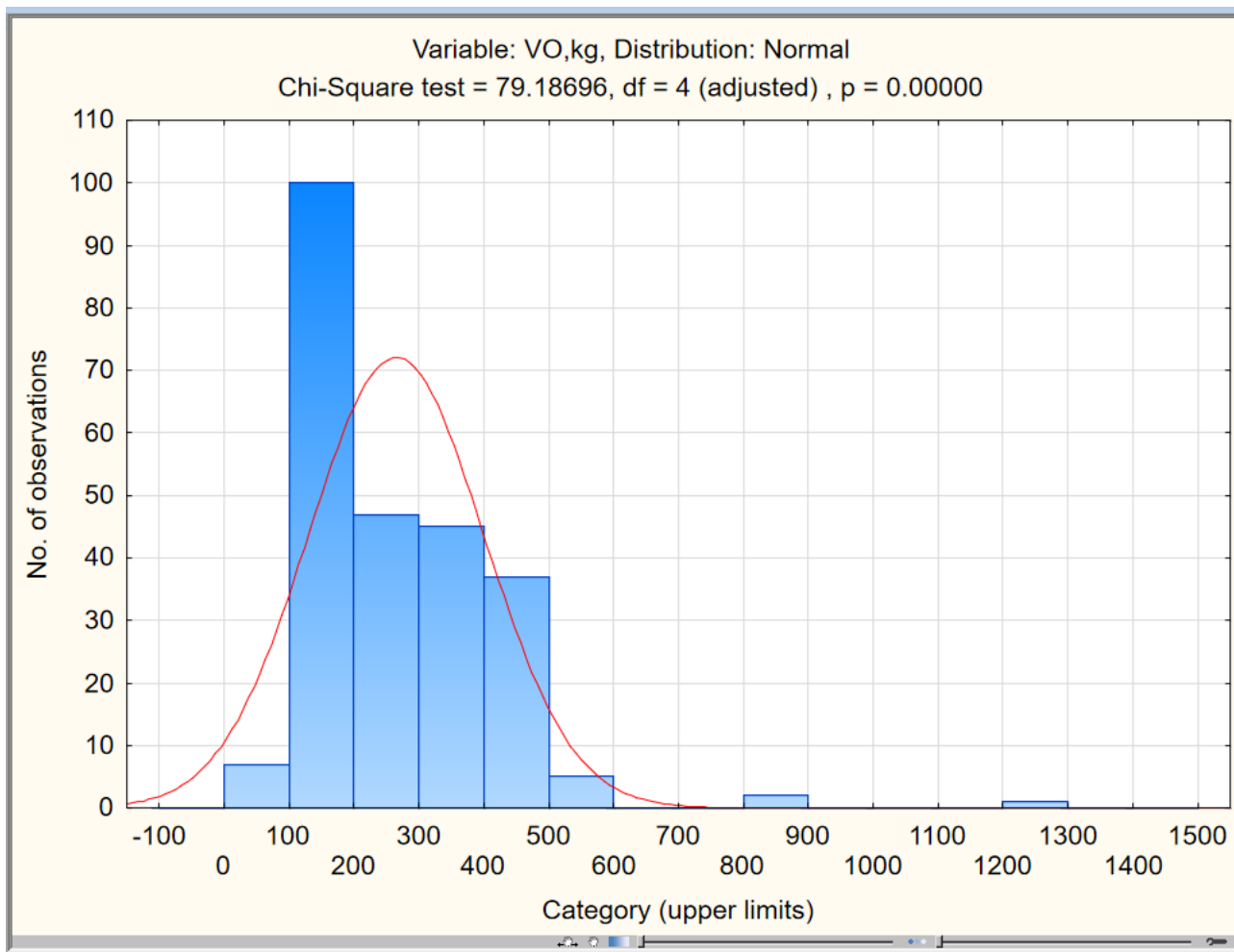


Рис. 3.1 Розподіл зворотніх відходів

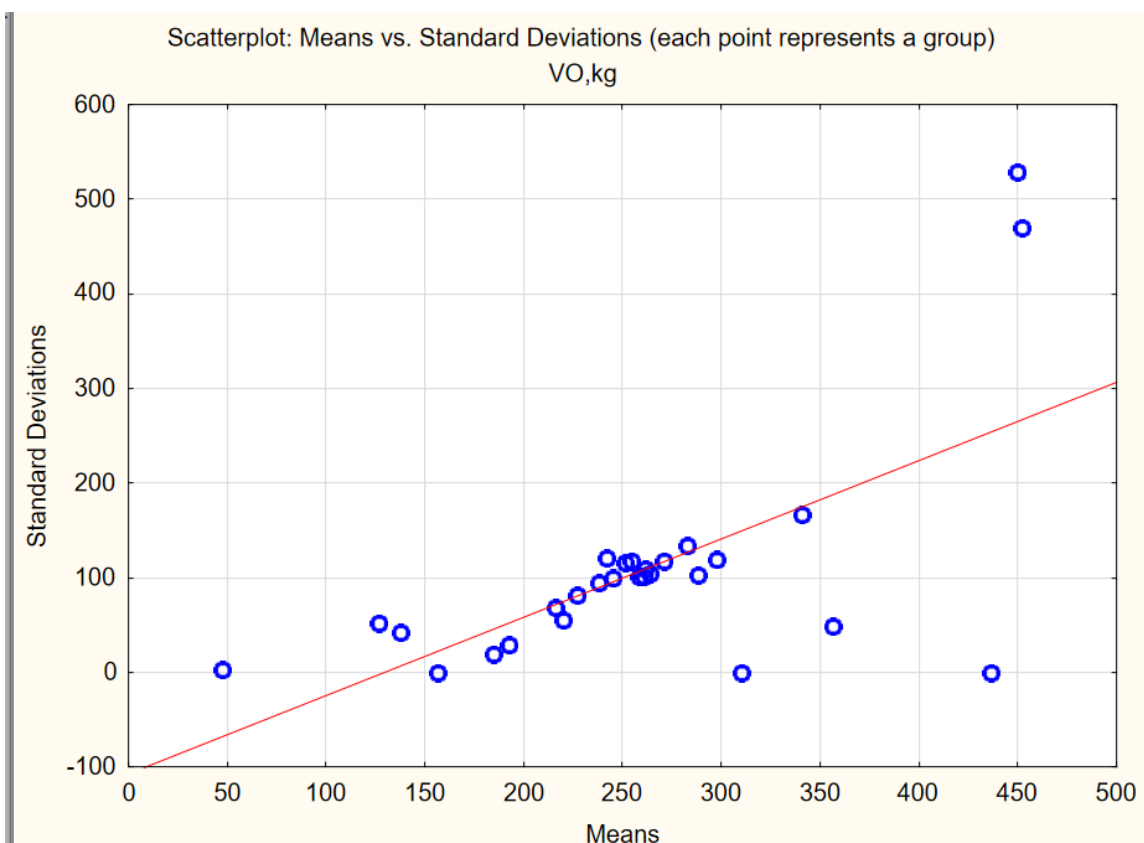
Variable: VO,kg, Distribution: Normal (Spreadsheet1) Chi-Square = 79.18696, df = 4 (adjusted) , p = 0.00000									
Upper Boundary	Observed Frequency	Cumulative Observed	Percent Observed	Cumul. % Observed	Expected Frequency	Cumulative Expected	Percent Expected	Cumul. % Expected	Observed-Expected
<= 0.00000	0	0	0.00000	0.0000	6.09152	6.0915	2.49653	2.4965	-6.0915
100.00000	7	7	2.86885	2.8689	21.03811	27.1296	8.62218	11.1187	-14.0381
200.00000	100	107	40.98361	43.8525	49.88701	77.0166	20.44550	31.5642	50.1130
300.00000	47	154	19.26230	63.1148	70.04749	147.0641	28.70799	60.2722	-23.0475
400.00000	45	199	18.44262	81.5574	58.26634	205.3305	23.87965	84.1518	-13.2663
500.00000	37	236	15.16393	96.7213	28.70491	234.0354	11.76431	95.9161	8.2951
600.00000	5	241	2.04918	98.7705	8.36783	242.4032	3.42944	99.3456	-3.3678
700.00000	0	241	0.00000	98.7705	1.44130	243.8445	0.59070	99.9363	-1.4413
800.00000	0	241	0.00000	98.7705	0.14641	243.9909	0.06001	99.9963	-0.1464
900.00000	2	243	0.81967	99.5902	0.00875	243.9997	0.00359	99.9999	1.9912
1000.00000	0	243	0.00000	99.5902	0.00031	244.0000	0.00013	100.0000	-0.0003
1100.00000	0	243	0.00000	99.5902	0.00001	244.0000	0.00000	100.0000	-0.0000
1200.00000	0	243	0.00000	99.5902	0.00000	244.0000	0.00000	100.0000	-0.0000
1300.00000	1	244	0.40984	100.0000	0.00000	244.0000	0.00000	100.0000	1.0000
1400.00000	0	244	0.00000	100.0000	0.00000	244.0000	0.00000	100.0000	0.0000
< Infinity	0	244	0.00000	100.0000	0.00000	244.0000	0.00000	100.0000	0.0000

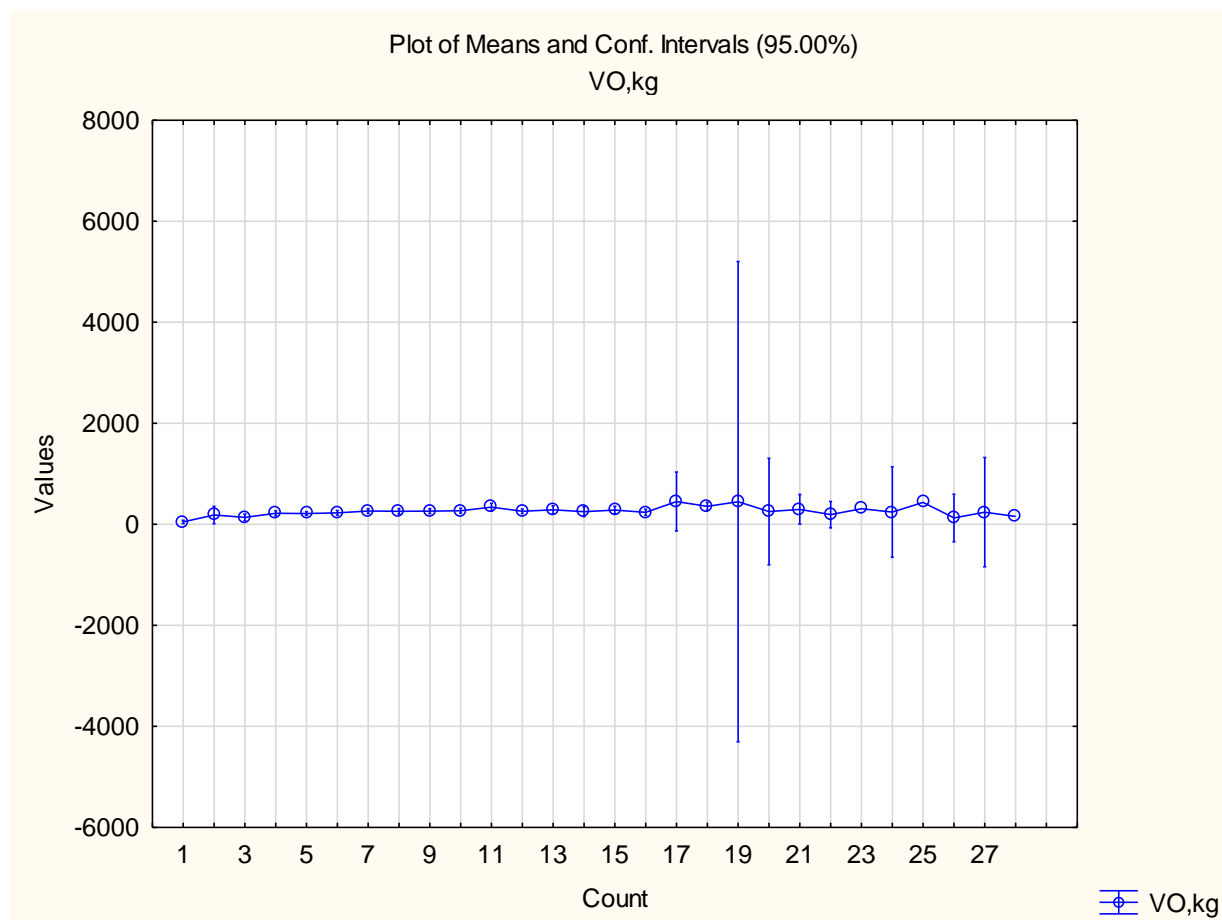
Рис 3.2 Показники розподілу зворотніх відходів

Перевіримо гіпотезу про залежність рівню зворотніх відходів від кількості поламок.

Analysis of Variance (Spreadsheet1) Marked effects are significant at p < .05000								
Variable	SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
VO,kg	760109.2	27	28152.19	3673611	216	17007.46	1.655285	0.026706

Як показую p-value ця залежність є статистично значущою, тобто чим більша кількість зупинок тим більші зворотні відходи на лінії.





Наступним кроком дослідимо, чи є залежність рівню зворотніх відходів від типу продукту. Як можна бачити на рис 3.3 зв'язок є та дуже статистично значущий.

Analysis of Variance (Spreadsheet1)								
Marked effects are significant at p < .05000								
Variable	SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
VO,kg	1054348	2	527173.9	3379373	241	14022.29	37.59541	0.000000

Рис.3.3 Показники дисперсійного аналізу зв'язку ЗВ від типу продукту

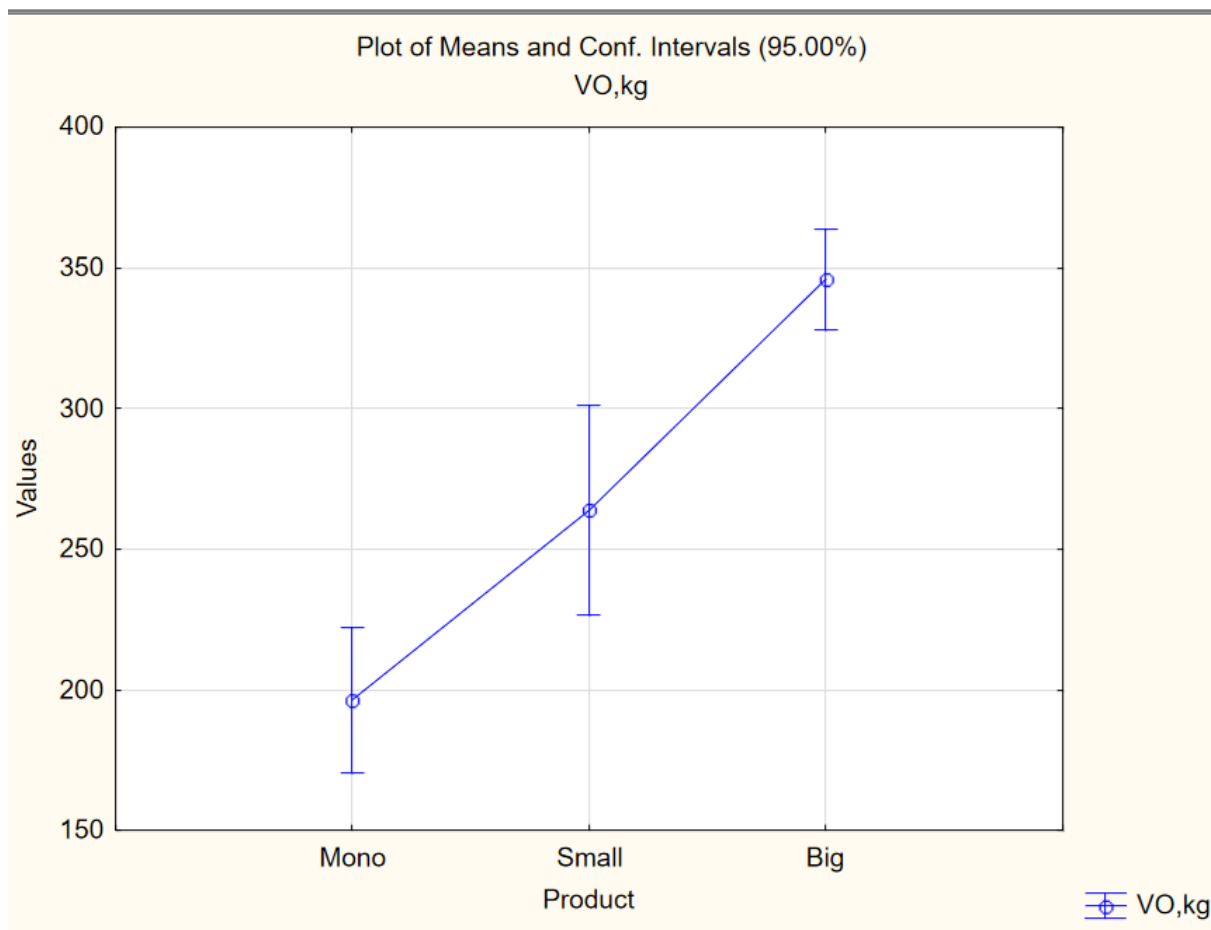


Рис.3.4 Графік середніх залежності ЗВ від типу продукту

Взявши до уваги, що ціна зворотніх відходів також відрізняється можна зробити висновок, що потрібна більш детально розглянути маржинальність різних продуктів, та окупованість цих витрат.

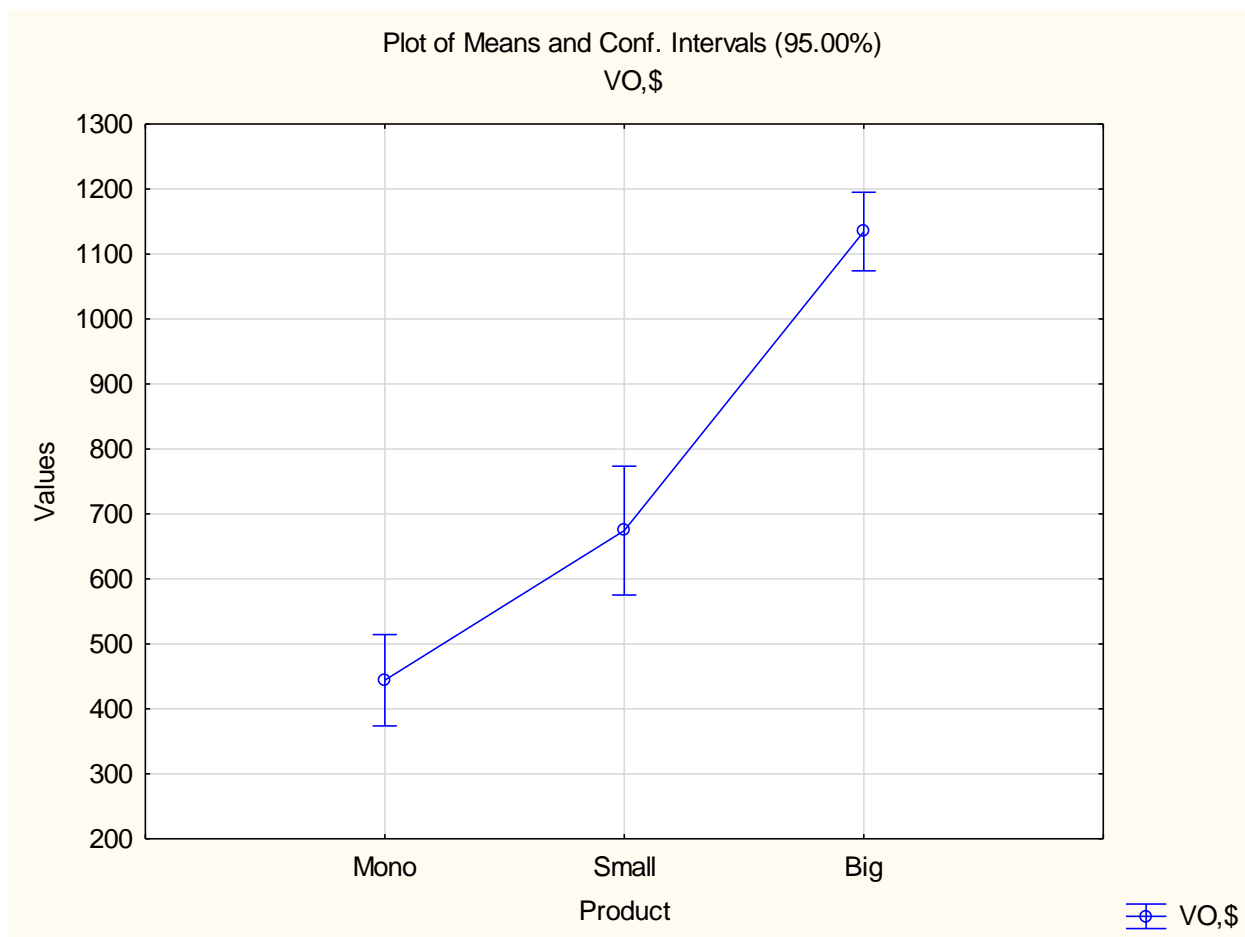


Рис 3.5 Зворотні відходи у валютному еквіваленті

Дослідимо наступну цікаву залежність. Це залежність загального часу простоїв від працюючої бригади. Тобто можна сказати від рівня кваліфікації. Якщо ця залежність буде статистично значущою, то можна робити вже дослідження за якою ціною буде вигідно навчати співпрацівників для зменшення рівня ЗВ та для збільшення прибутку. На рисунку 3.6. ми бачимо, що p -value дуже близько до 0.05 тобто це означає що є залежність професіоналізму та часу зупинок. Детальніше розглянемо на рис 3.7

Analysis of Variance (Spreadsheet1)								
Marked effects are significant at $p < .05000$								
Variable	SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
Pause, min	48099.56	12	4008.297	213918.3	96	2228.316	1.798801	0.058940

Рис 3.6 Результати дисперсійного аналізу хв. Простою від кваліфікації

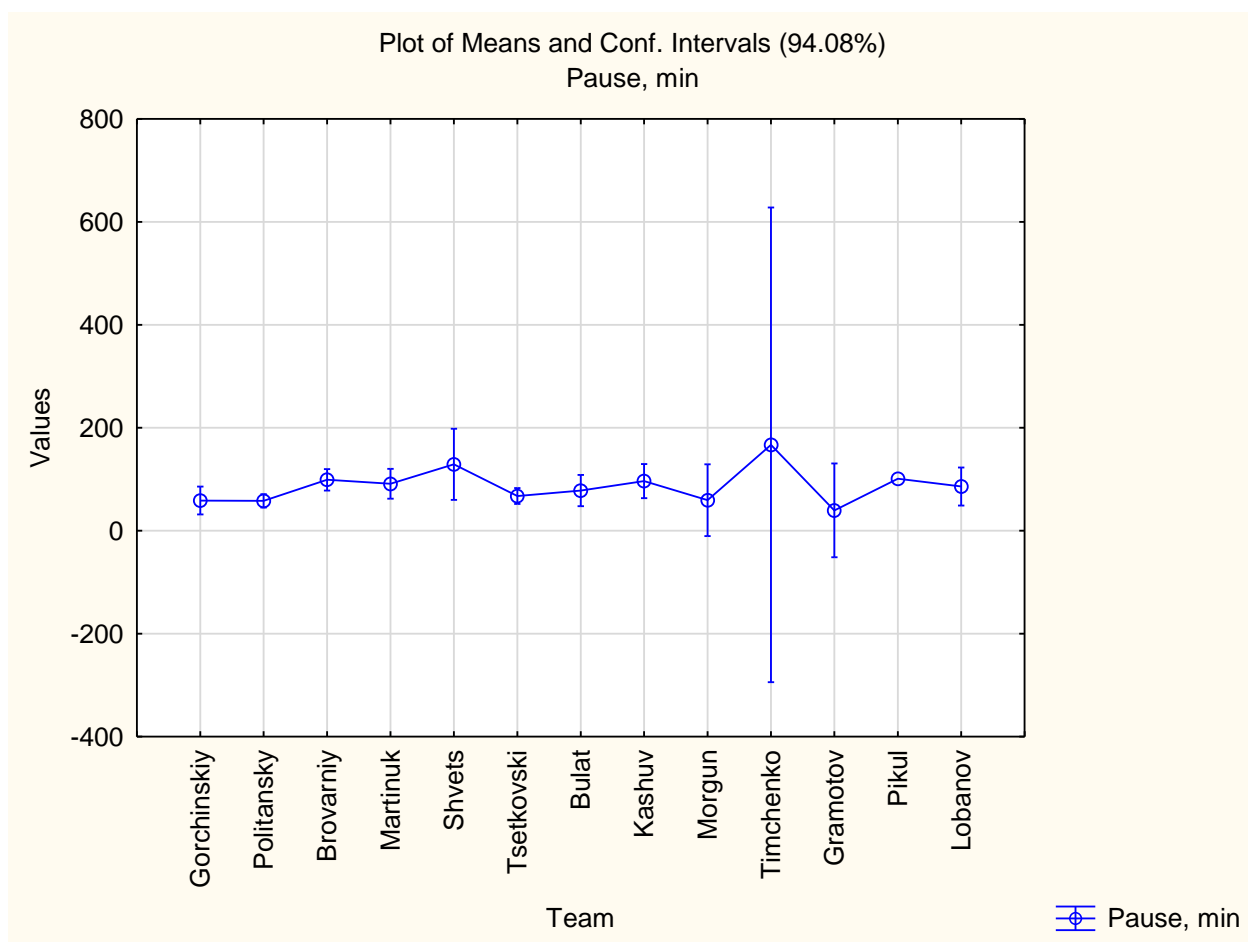


Рис 3.7. Залежність хв простою від кваліфікації бригади

Зупинимось детальніше на робітниках. Кластеризуємо співробітників в залежності від стажу роботи. У першу категорію віднесемо працівників зі стажем роботи 15 і більше років. У другу від 8 до 15, а третю до 8 років. Отримаємо результати на рисунку 3.8

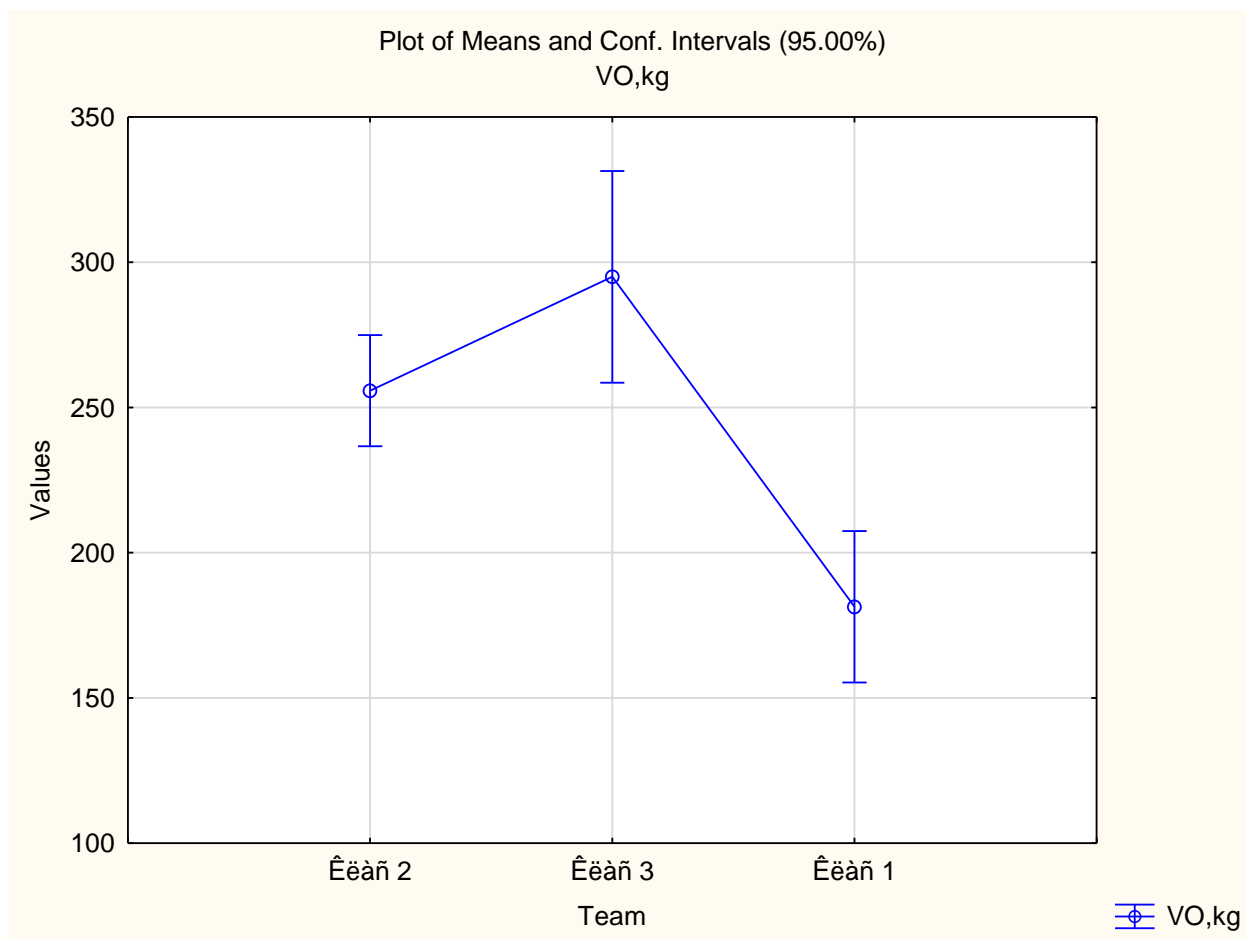


Рисунок 3.8. Згруповані співробітники за стажем роботи.

Наступним кроком є перевірка гіпотези про вплив постачальника. Нажаль на цій лінії є тільки три постачальника, тому дослідження буде не дуже коректним. Найяскравіші результати ми побачимо на наступній лінії, де постачальник грає важливу роль.

Univariate Tests of Significance for VO,kg (Spreadsheet1)					
Sigma-restricted parameterization					
Effective hypothesis decomposition					
Effect	SS	Degr. of Freedom	MS	F	p
Intercept	1058694	1	1058694	60.88298	0.000000
Saler	109887	2	54943	3.15966	0.044218
Team	180153	2	90077	5.18009	0.006276
Error	4155971	239	17389		

Рисунок 3.9 Показники впливу постачальника та бригади

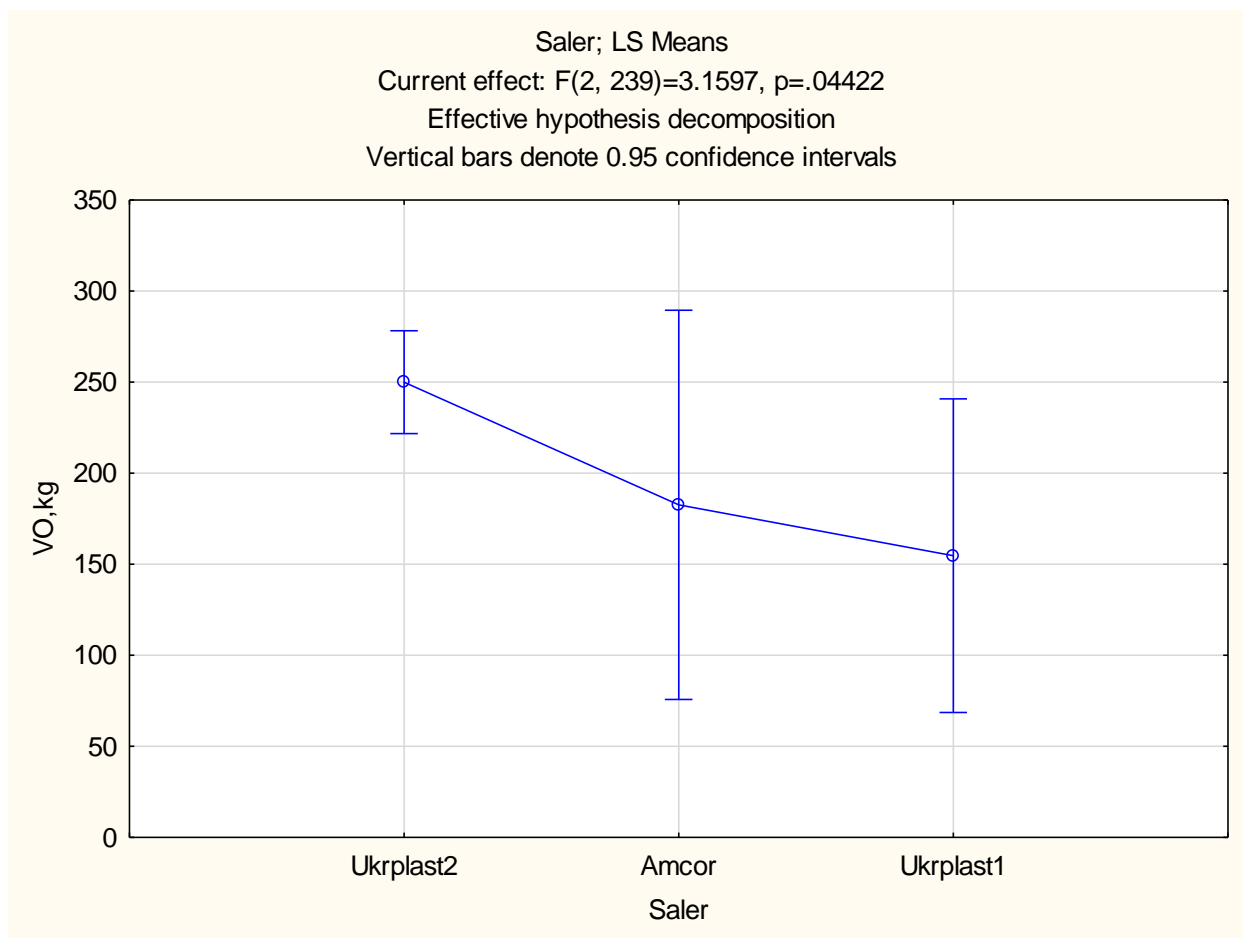


Рисунок 3.10 Графики середніх постачальників

Як ми бачимо хоча p -value показує статистичну значущість, але це не так явно як ми це побачимо на прикладі іншої лінії.

Друга лінія шоколаду така сама, але більш проблемна. Зробимо ті ж самі аналізи та порівняємо висновки.

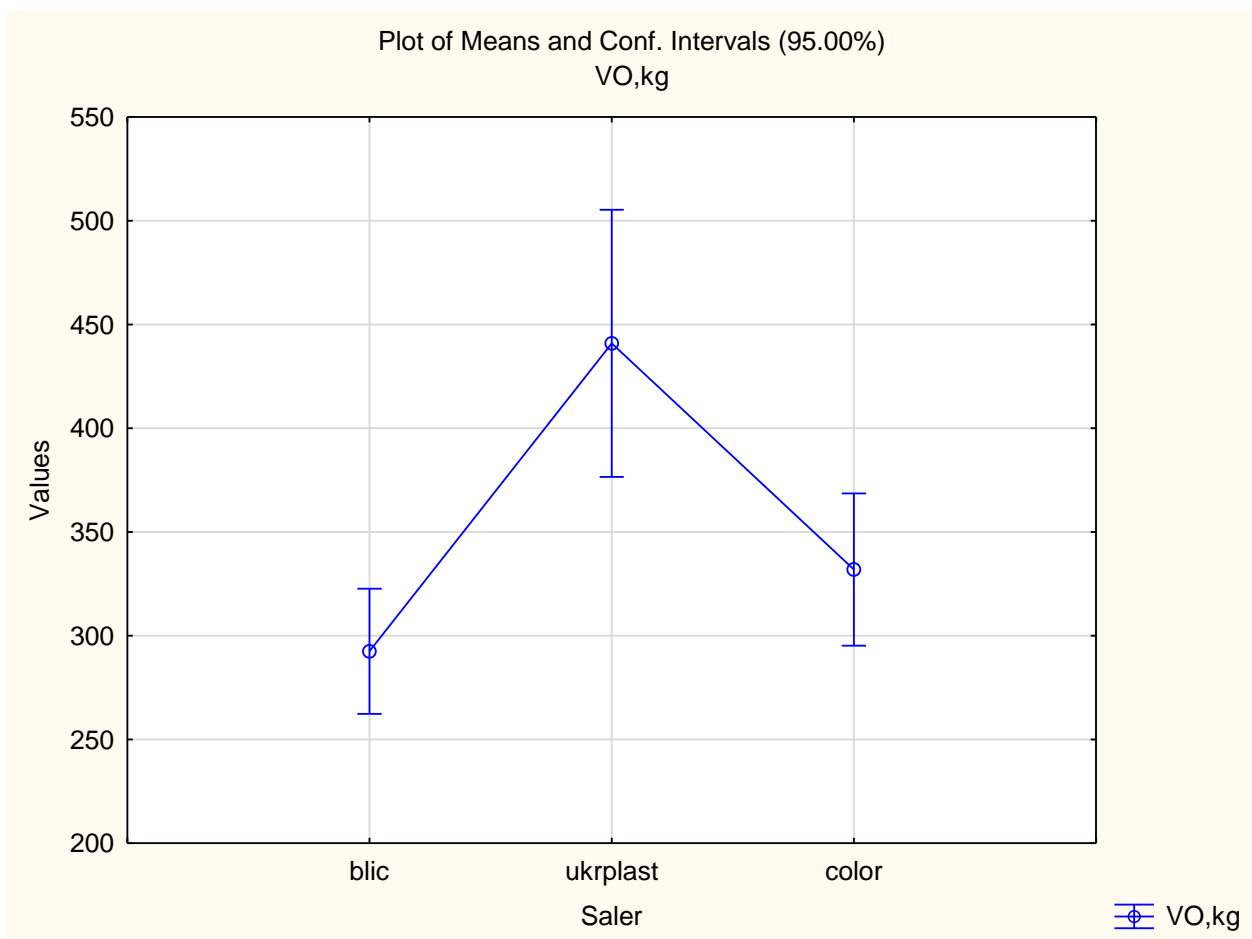


Рисунок 3.11 Аналіз постачальника етикету

Analysis of Variance (Spreadsheet1)								
Marked effects are significant at p < .05000								
/variable	SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
VO,kg	1781554	2	890776.8	923387.9	106	8711.207	102.2564	0.000000

Рисунок 3.12 Аналіз постачальника етикету

Спостерігаємо таку ж саму залежність ЗВ від типу продукту. Можна зробити висновки, що додаткова оцінка маржинальності є вкрай необхідною для більш якісного аналізу.

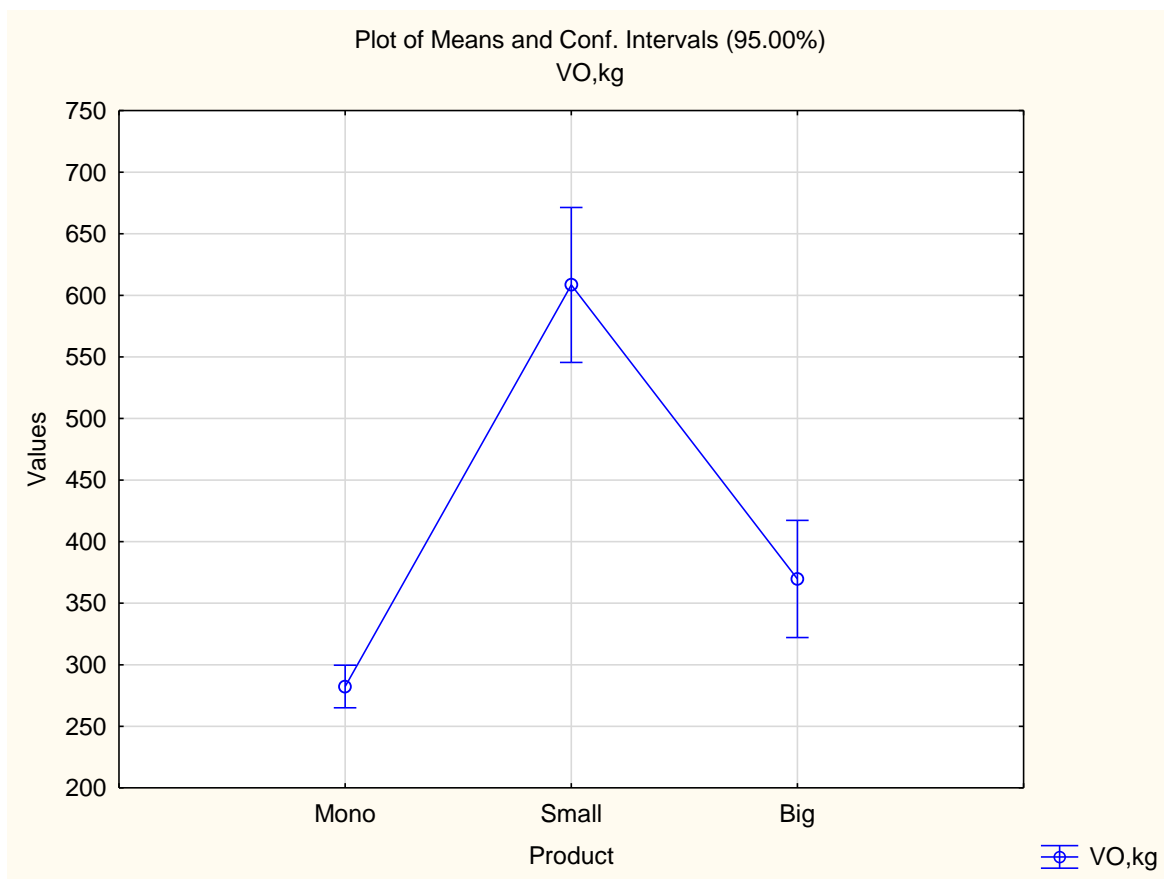


Рисунок 3.13 Рівень ЗВ від типу продукту

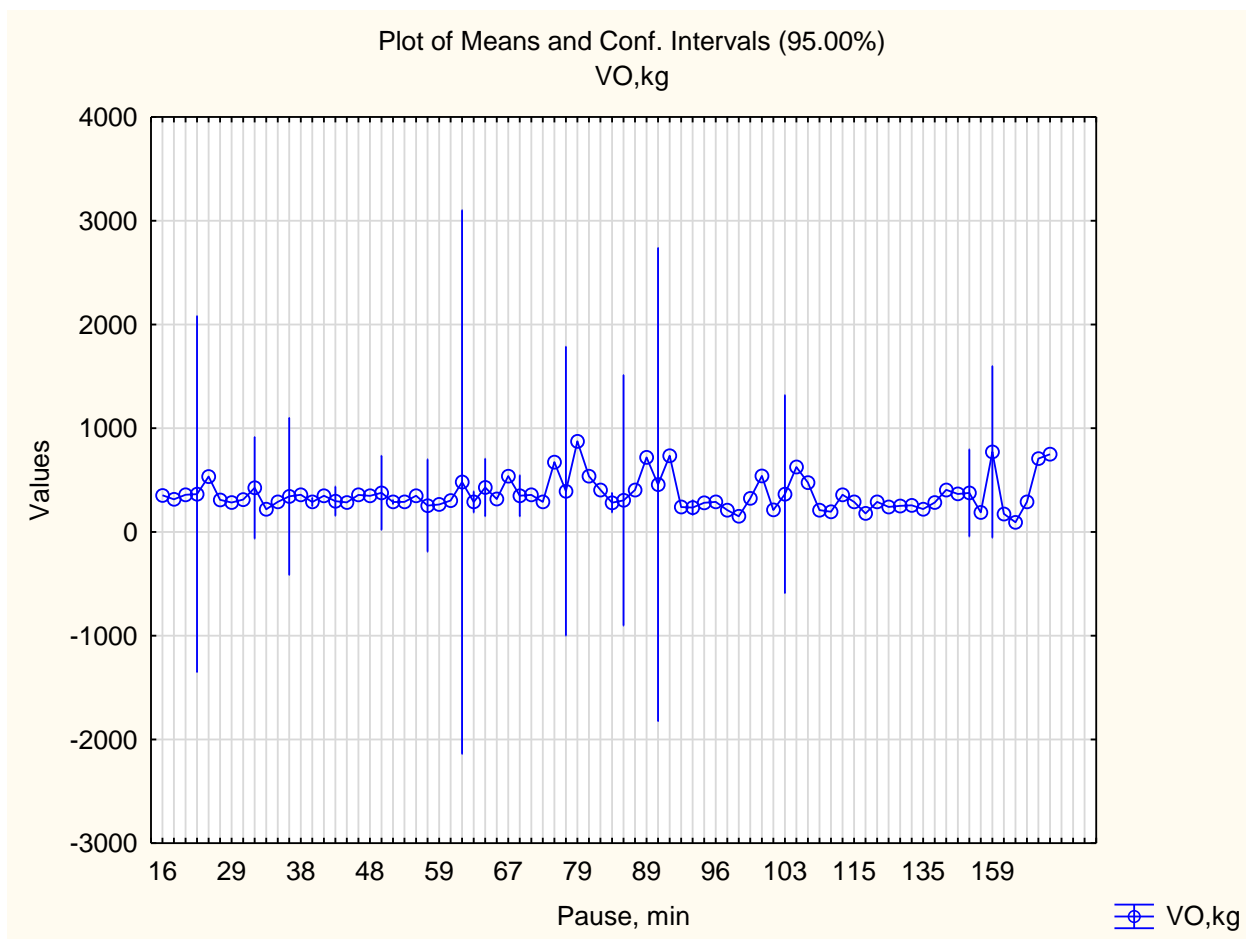


Рисунок 3.14 Залежність ЗВ від тривалості зупинок

3.2 Побудова архітектури програми по підтримці прийняття рішень

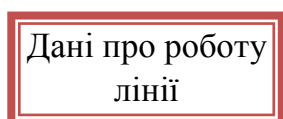
У цьому розділі ми розглянемо систему підтримки прийняття рішень для прогнозування та прийняття рішення щодо фінансових вкладень для мінімізації витрат на виробничій лінії.

Порядок роботи з СППР:

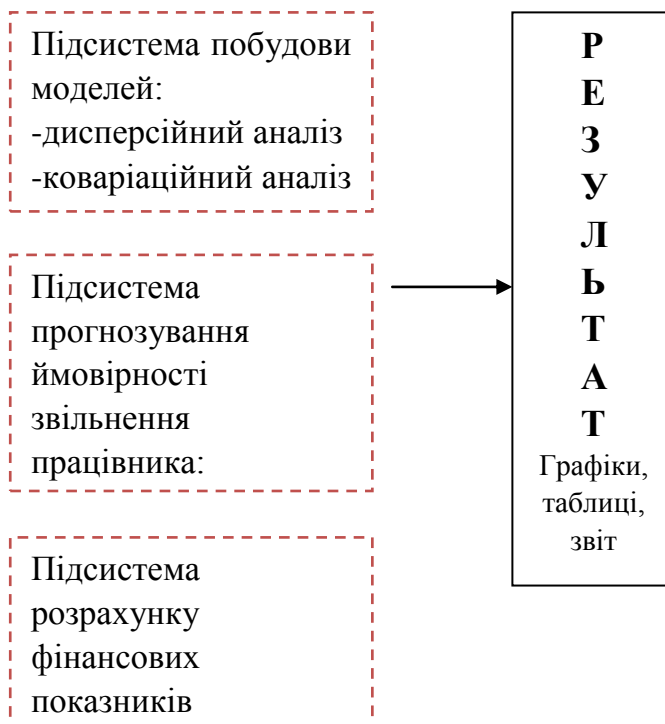
- 1) Обираємо досліджуєму лінію. Наприклад: Winkler 1, Winkler 2 чи іншу з списку.
- 2) Обираємо період та завантажуюмо дані із системи Scada про швидкість відливання.
- 3) Система трансформує дані про роботу лінії у інший формат, а саме дата/зміна/загальний час простоїв/кількість простоїв

- 4) Обираємо період та завантажуюмо дані із системи IT Enterprise щодо виробітку та кількості зворотніх відходів, а також цін на етикет по кожному постачальнику.
- 5) СППР перетворює у потрібний формат та перевіряє сопоставимість з даними по роботі лінії. Якщо знайдена невідповідність видається повідомлення та пропонується вибір рішення чи пропустити ці дані, чи заповнити попереднім значенням.
- 6) Обираємо період та завантажуюмо дані із системи Roshen Horses по виходу конкретної людини у ці зміни, та дані по спрацюванню етикету та його наєжності постачальнику
- 7) Якщо є необхідність то додаємо базову матрицю кваліфікацій співробітників. В якій вказано всі дані про співробітників. При знаходженні пропусків видається повідомлення.
- 8) Проводиться аналіз введених даних. Виводяться показники p-value по кожному обраному фактору.
- 9) Далі виводяться існуючі фінансові показники витрат. Та гранично допустимі вартості фінансових покращень по кожному з факторів. Тобто шляхи покращення для кожного з факторів. Наприклад, аналізуючи залежність постачальника та його ціни пропонується купляти більш дорогий етикет для того щоб зекономити зворотні відходи. Чт пропонується навчити конкретну групу людей, для того щоб перевести в інший кластер і тим самим також зменшити зворотні відходи та час простою лінії.
- 10) Після обраних рішень проводиться прогнозування фінансових показників з урахуванням обраних змін.

На рисунку 3.2.1 наведена структура розробленої СППР. Як можна побачити з цієї структури СППР представляє собою широкий комплекс засобів для аналізу та обробки даних.



Блок аналізу та прогнозування



На рисунку 3.2.1 Структура системи підтримки прийняття рішень

Усі дані можна завантажити з різних систем і баз за допомогою пристроїв вводу-виведення СППР. Для цього підсистема функціонально зв'язана з навними базами даних фірми. Підсистема інтерфейсу користувача призначена для здійснення зв'язку між користувачами СППР та внутрішніми елементами системи і забезпечує ввід та вивід інформації, а також надає досуп до зовнішніх запам'ятовуючих пристроїв ПЕОМ. Інтерфейс дозволяє користувачу вводити дані, інформацію про всі види продукції, команди, параметри і запити в систему та одержувати вихідну інформацію в зручному вигляді. Також він дозволяє перевіряти нові сформовані системою дані. В зручному вигляді легше знаходити можливі помилки.

Основною складовою підсистеми збереження інформації є база даних, яка призначена для накопичення статистичних даних які було завантажено у систмеу, з метою їх подальшого аналізу, обробки та використання.

Підсистема обробки інформації потрібна для перевірки даних, наявності можливих пропусків чи несумісностей. Наприклад, якщо дані про роботу лінії автоматично зібрані з системи свідчать про те, що лінія працювала у той день, а дані про виробіток і зворотні відходи свідчать про простій лінії, то видається звіт цих помилок і несумісностей, для того щоб аналітик міг вручну проаналізувати та зробити висновки чи це помилка системна, чи випадкова.

Блок аналізу та прогнозування складається з чотирьох підсистем: підсистема побудови моделей, підсистема розрахунку фінансових показників та підсистема прогнозування ймовірності звільнення працівників.

Підсистема побудови моделі реалізована за допомогою двох математичних методів: дисперсійного аналізу, та коваріаційного аналізу.

Підсистема прогнозування ймовірності звільнення працівників обчислює статистичні показники побудованої моделі ймовірностей.

Підсистема розрахунку фінансових показників обраховує вигідність від одних чи інших вкладень, обчислює амортизацію та окупованість продукту.

Підсистема виведення результатів представляє собою набір графіків, таблиць та звітів для прийняття рішення експертом. Представлення результатів прогнозування та критеріїв оцінювання моделі дає змогу зробити висновки щодо можливості використання отриманої моделі для прогнозування.

Для створення СППР застосовувались технології статистичної мови R.

На теперішній час більшість людей користується різними операційними системами тому для того щоб забезпечити легкий доступ до СППР вона була написана у вигляді веб додатку для браузера. Це спрощує роботу користувача, робить зайвим запам'ятовування великої кількості команд для виконання певних дій та подає отримані результати у зручному вигляді.

3.3 Основні технічні вимоги для коректної роботи програми

Для роботи програмного продукту необхідна наявність персонального комп'ютера з наступними мінімальними характеристиками:

- а) будь-яка операційна система з встановленим браузером та Adobe Flash player;
- б) тактова частота процесора 1 ГГц;
- в) оперативна пам'ять розміром 512 Мбайт;
- г) вільний дисковий простір: 5 Мбайт для розміщення виконавчого файлу, вхідних даних і результатів роботи;
- д) клавіатура та комп'ютерна мишка;
- е) монітор з розподільчою здатністю 1024×768;

3.4 Інструкція з експлуатації програмного продукту

Робота з усіма елементами інтерфейсу є стандартною для програмного забезпечення, що працює на платформі будь-якого браузера. Усі можливі некоректні введення даних обробляються системою та попереджують користувача інформаційними повідомленнями.

Основний робочий екран СППР має структуру наведену на рисунку 3.4.1

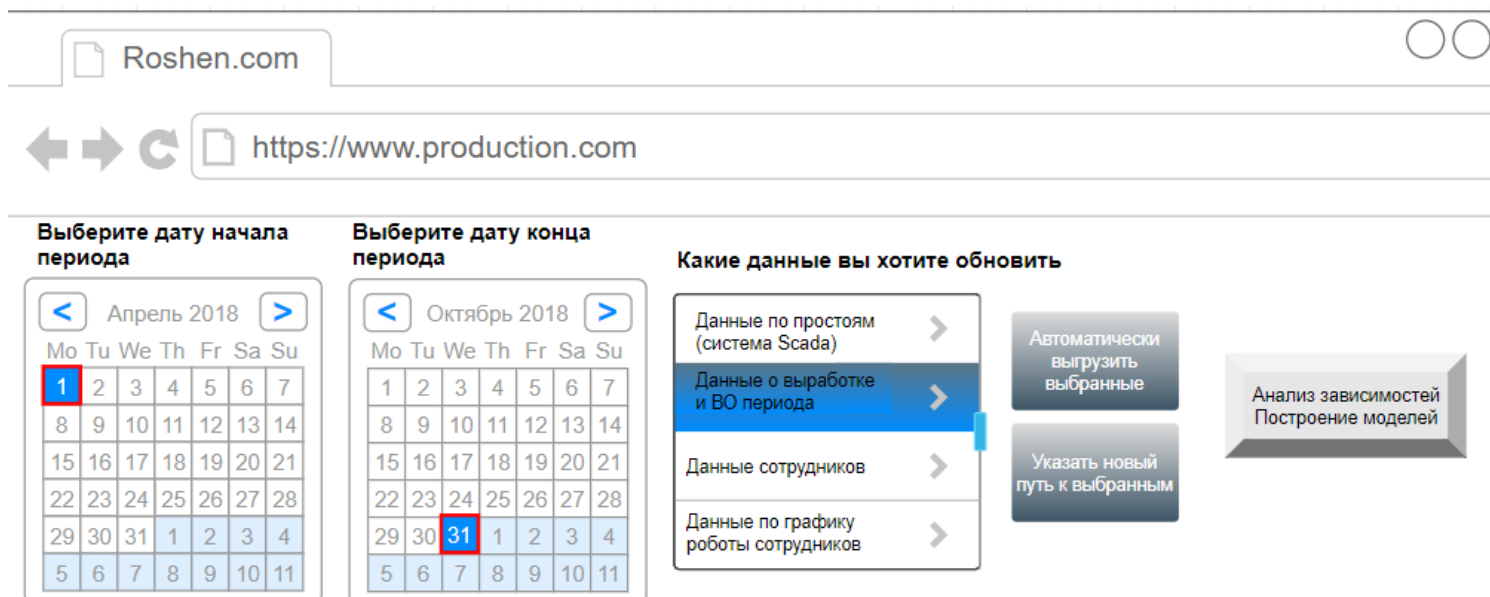


Рисунок 3.4.1 Головне вікно СППР

Робота з програмою починається з того, що потрібно вибрати період який ми хочемо аналізувати та на основі якого буде прийматися рішення. Це зроблено для того щоб аналітик міг спеціально виключати дати, які з якихось причин непригодні для аналізу. Наприклад, в ті дати були техзапуски, чи вимкнення електроенергії.

Потім пропонується обрати які саме данні ви хочете завантажити для аналізу. І у вас є дві можливості: або завантажити з прописаних систем (Scada, IT Enterprise, Horses) або завантажити вже готовий сводний вихідний файл з комп'ютера.

Якщо ви обрали завантажувати з баз даних то натискаємо першу кнопку. Можливі два випадки : якщо все завантажилось успішно (Рис 3.4.2), та другий коли є якась помилка в даних (рис. 3.4.3). У другому випадку, можна відкрити звіт з помилками та для кожної окремо або заповнити дані, або пропустити ці рядки даних. Це показано на рисунку 3.4.4

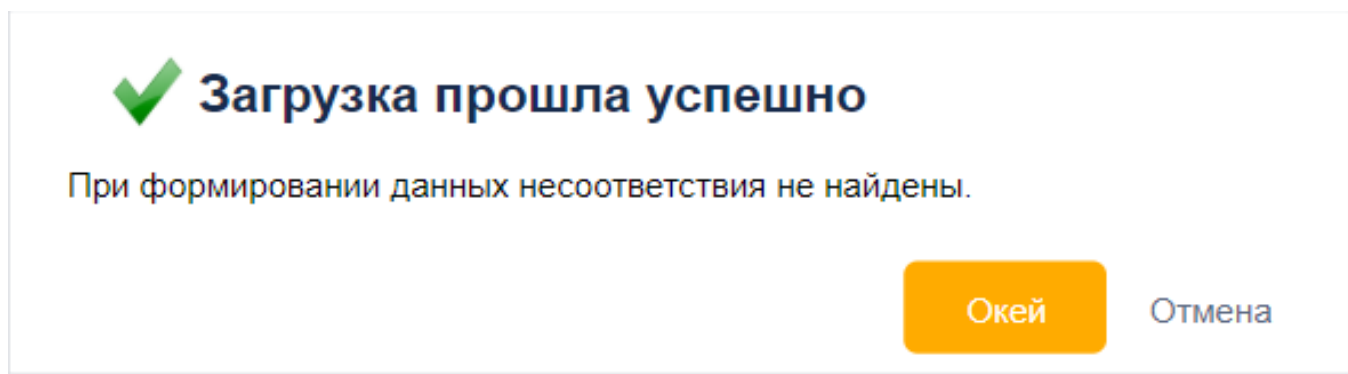


Рисунок 3.4.2 Повідомлення про успішне завантаження даних

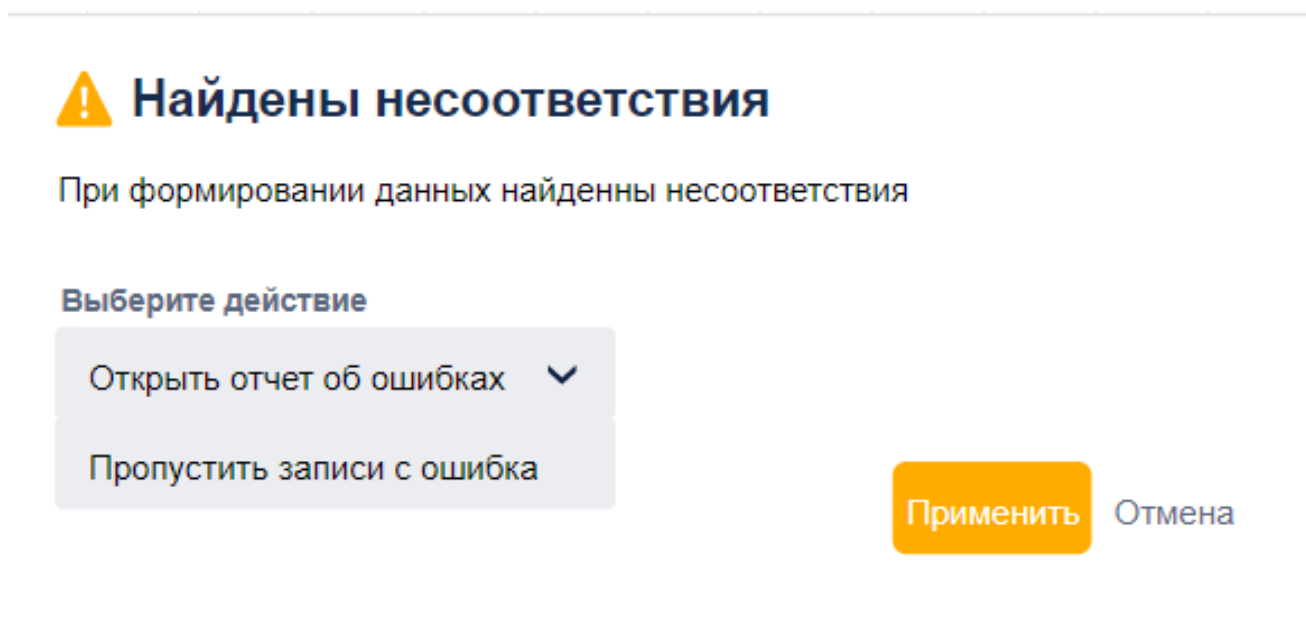


Рисунок 3.4.3 Повідомлення про знайдені помилки при завантаженні даних

	1	2	3	4	5	6	7	8	9	10	11
	Дата	Смена	Простои, мин	Работа, мин	Кво	Тип	Поставщик	Бригада	ВО, кг	ВО, \$	Выработка
⊗	1 23-Apr-18	3	158	322	15	Small	Ukrplast2		255	659.95	834
	2 24-Apr-18	1	115	364	14	Big			174	578.62	873
⊗	3 24-Apr-18	2		446	10	Big	Ukrplast2	Politansky	280	931.11	116
⊗	4 24-Apr-18	3	131		7	Big	Ukrplast2	Kashuv	295.2		989
	5 25-Apr-18	1	23	456	8	Small		Gorchinskiy	212	601.67	1393
	6 25-Apr-18	2	105	375	4		Ukrplast2	Prilutsky	148.6	421.74	117

Рисунок 3.4.4 Звіт з помилками та редакцією

Якщо на етапі повідомлення про знаходження помилки обрати «Пропустить записи с ошибками» то всі рядки які є неповними будуть видалені з ітогової таблиці.

Червоні хрестики у звіті зліва означають, що аналітик виключив ці рядки з розрахунку, якщо він заповняє якесь поле, то це одразу записується в базу.

Якщо на першому кроці користувач вирішив не завантажувати окремі дані з різних систем, а сам якось їх сформував, то є можливість відкрити та аналізувати напряму цей файл. Виглядає це стандартно, показано на рисунку 3.4.5.

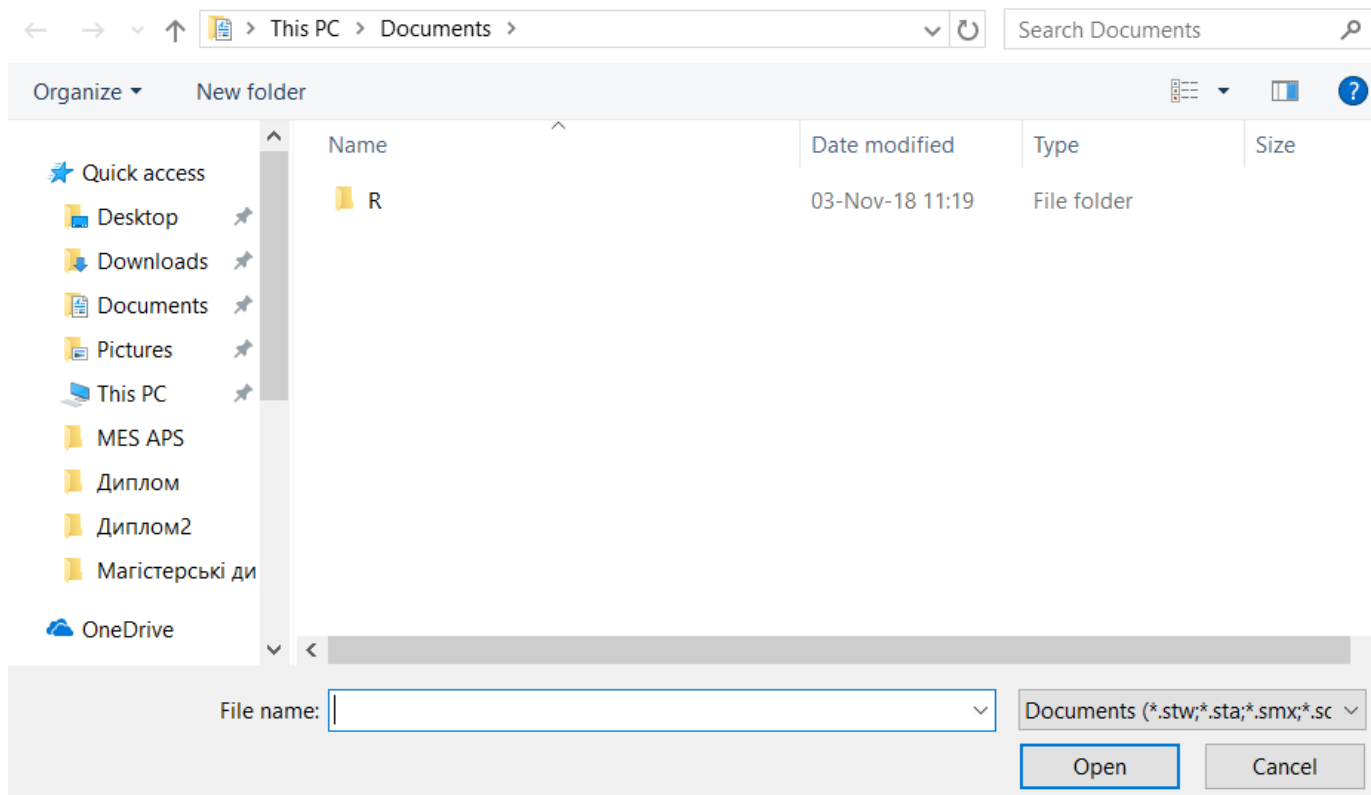


Рисунок 3.4.5 Вибір файлу для завантаження

Дані можуть бути в форматах – xls/xlsx/csv. Для підготовки даних до аналізу найкраще застосовувати програму Microsoft Excel. Дані можуть приймати як числові, так і строкові значення. Для завантаження даних необхідно перейти по головному меню по натисканню кнопки «Указать новый путь к выбранным категориям». Після чого відкриється стандартне діалогове вікно операційної системи вибору директорії з файлами (рис. 3.4.5), в якому слід обрати формат необхідного файлу з даними, та безпосередньо сам файл даних.

Після вибору файлу даних відкриється вікно з попереднім переглядом даних про обраний файл (рис. 3.4.6), а саме буде відображено: формат файлу та

шлях його розташування в пам'яті комп'ютера. У цьому вікні необхідно вказати чи містить набір даних перший рядок з назвами змінних, поставивши або прибравши відповідний прапорець. Але варто зазначити, що для правильного аналізу порядок стовпчиків повинен бути визначений програмою.

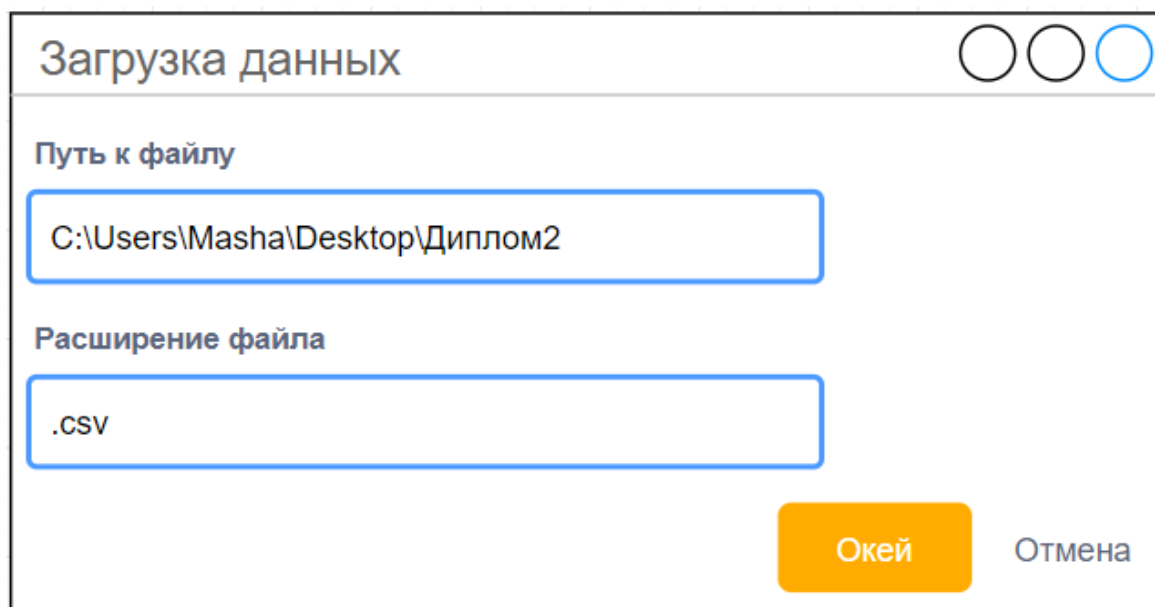


Рисунок 3.4.6 Вікно завантаження даних. Перевірка даних

Після натискання «Окей» файл буде оброблено та завантажено до системи. Також дані файлу відображаються на головному екрані програми, що зображено на рисунку 3.4.7. Це зроблено для того щоб аналітик міг ще раз передивитись завантажені дані, також доступна функція корегування. Якщо хоче щось підправити чи змінити можна це зробити одразу у головному вікні.

Зверніть увагу, що при редагуванні даних у діалоговому вікні, вони не зберігаються автоматично до ісходного файлу. Це на данному етапі є недороботкою програми. Але якщо після редагування перейти по кнопках «Перейти к окну анализа данных» або «Перейти к окну финансовых показателей» то в роботу возьметься именно последняя отредактированная версия.

Выберите дату начала периода

Выберите дату конца периода

Какие данные вы хотите обновить

Данные по простоям (система Scada)

Данные о выработке и ВО периода

Данные сотрудников

Данные по графику работы сотрудников

Автоматически выгрузить выбранные

Указать новый путь к выбранным

Анализ зависимостей
Построение моделей

	1 Дата	2 Смена	3 Простой, мин	4 Работа, мин	5 Кво	6 Тип	7 Поставщик	8 Бригада	9 ВО, кг	10 ВО, \$	11 Выработка, шт
1	03-Apr-18	1	133	346	26	Mono	Ukrplast2	Gorchinskiy	164	382.51	11460.96
2	03-Apr-18	2	38	442	12	Mono	Ukrplast2	Politansky	172	401.17	11748.24
3	03-Apr-18	3	122	358	9	Small	Ukrplast2	Brovarniy	189.6	538.1	13932
4	04-Apr-18	1	49	430	15	Small	Ukrplast2	Martinuk	174.4	494.96	11707.2
5	04-Apr-18	2	57	423	9	Mono	Ukrplast2	Shvets	171	397.12	11793.6
6	19-Apr-18	2	123	357	11	Big	Ukrplast2	Brovarniy	316.6	1050.44	8772.57
7	19-Apr-18	3	67	413	9	Mono	Ukrplast2	Tsetkovski	212	492.34	14044.59
8	20-Apr-18	1	75	404	10	Small	Ukrplast2	Politansky	290	799.61	11111.58
9	20-Apr-18	2	92	388	6	Small	Ukrplast2	Brovarniy	335.7	925.62	11647.26
10	20-Apr-18	3	177	303	10	Big	Ukrplast2	Gorchinskiy	452	1479.72	8764.47
11	23-Apr-18	2	98	382	8	Mono	Ukrplast2	Bulat	177	412.83	11779.02
12	23-Apr-18	3	158	322	15	Small	Ukrplast2	Brovarniy	255	659.95	8341.11
13	24-Apr-18	1	115	364	14	Big	Ukrplast2	Gorchinskiy	174	578.62	8730.72
14	24-Apr-18	2	34	446	10	Big	Ukrplast2	Politansky	280	931.11	11620.8
15	24-Apr-18	3	131	349	7	Big	Ukrplast2	Kashuv	295.2	981.66	9898.74
16	25-Apr-18	1	23	456	8	Small	Ukrplast2	Gorchinskiy	212	601.67	13932.54
17	25-Apr-18	2	105	375	4	Small	Ukrplast2	Prilutsky	148.6	421.74	11763.9
18	25-Apr-18	3	138	342	3	Small	Ukrplast2	Brovarniy	140.5	359.69	11638.62
19	10-May-18	1	238	241	13	Small	Ukrplast2	Kashuv	231.2	598.36	6237.9
20	10-May-18	2	80	400	5	Mono	Ukrplast2	Tsetkovski	171	412.96	11461.41
21	10-May-18	3	99	381	3	Mono	Ukrplast2	Politansky	163	393.64	11832.93
22	11-May-18	1	18	461	5	Mono	Ukrplast2	Brovarniy	212	492.34	13932.45
23	11-May-18	2	121	359	12	Mono	Ukrplast2	Tsetkovski	150	348.35	11720.61
24	11-May-18	3	146	334	12	Big	Ukrplast2	Goroshko	291.6	954.62	8650.98
25	12-May-18	1	16	463	5	Big	Ukrplast2	Kashuv	421	1396.83	11620.71

Перейти к описанию анализа данных

Перейти к описанию финансовых показателей

Рисунок 3.4.7 Головне вікно після завантаження даних

Розглянемо більш детально нове вікно для аналізу даних. В ньому представлено три відділу. Перший відділ наданий для аналізу статистики по звільненню працівників, наводяться діаграми та статистика по звільненню. На основі цих даних пропонується побудувати логістичну регресію, яка буде прогнозувати чи звільниться працівник чи ні.

У

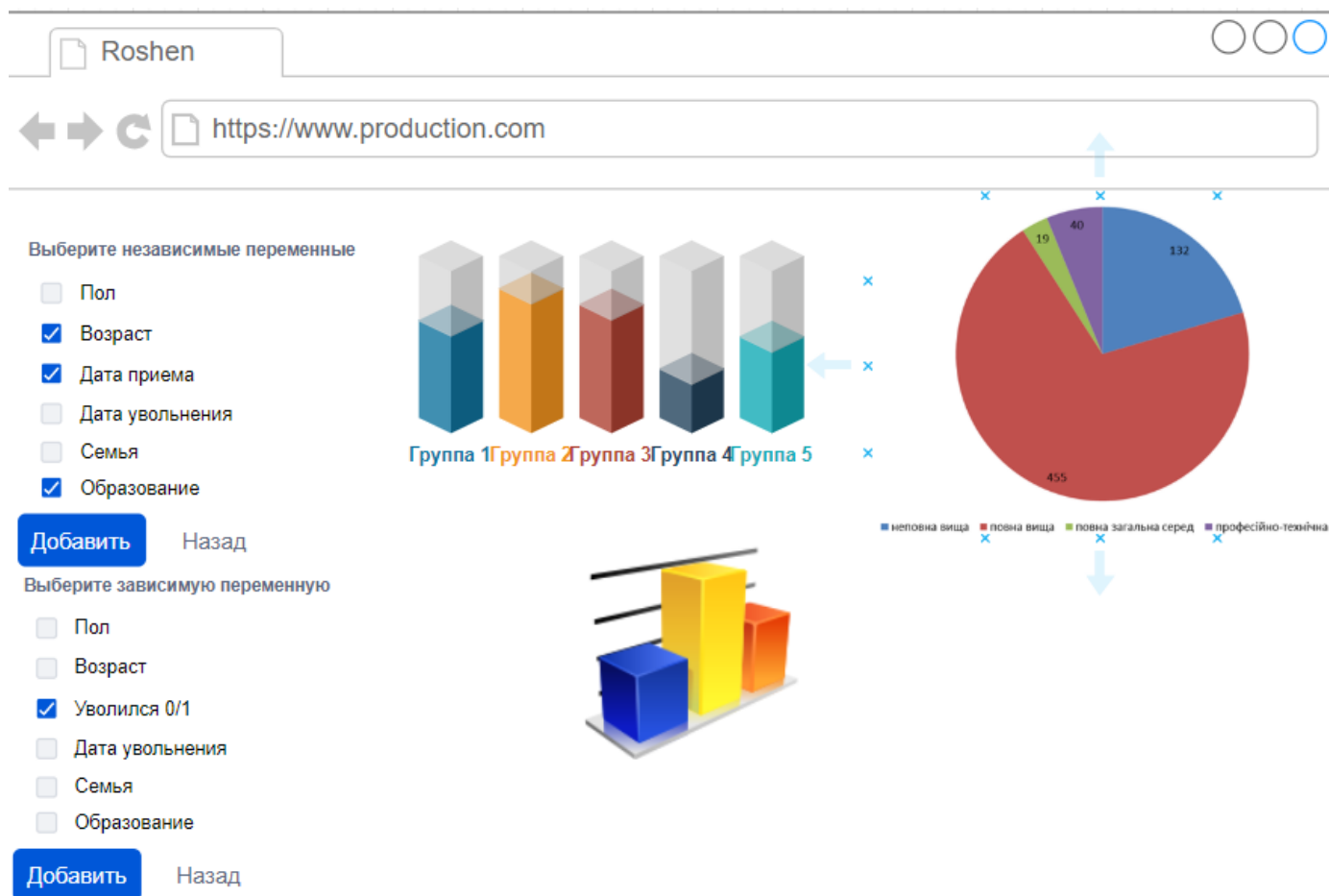


Рисунок 3.4.8 Вікно для аналізу працівників та статистики по звільненню.

У другому відділі пропонується зробити дисперсійний аналіз та коваріаційний аналіз. Пропонується за допомогою галочок визначати які фактори ви обираєте як предиктори які як фактори.

Наступним кроком будується сама регресійна модель рис 3.4.9

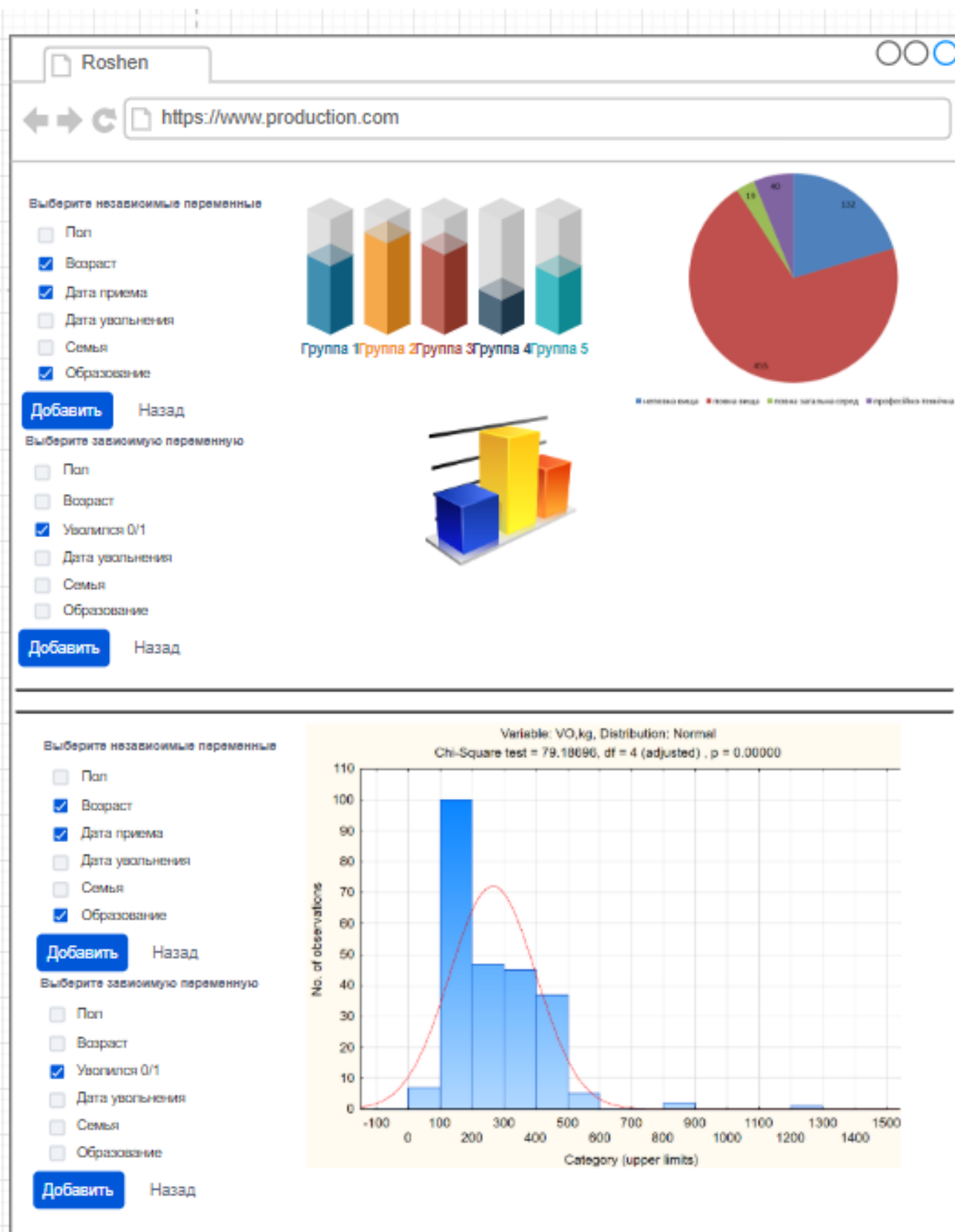


Рисунок 3.4.9 Вікно для дисперсійного і коваріаційного аналізу

Якщо прокрутити нище пропонується ряд аналізів та методів для прогнозування величин. Після цього можна переходити у вікно з фінансовими показниками у якому вже моделювати зміни цін та проведення навчання.

3.5 Використання результатів програми

Після обробки даних та проведення аналізу на виробничій лінії Winkler 1 було отримано наступні результати у таблиці 3.5.1:

	SS	Степеней свободы	MS	F	p
Свободный член	4744328	1	4744238	494,3307	0,000000
Квалификация сотрудника	305809	4	77202	8,0442	0,000003
Поставщик этикета	106856	4	27380	2,8566	0,534654
Тип продукта	45675	3	17083	43,3836	0,0478083
Количество остановок	456463	4	26574	2,1164	0,0431546

Таблица 3.5.1 Результаты анализа линии Winkler 1

З таблиці легко бачити, що у допустимій мірі впливають кваліфікація співробітника, тип продукта та кількість зупинок. А от постачальник етикету не впливає на кінцевий результат. Це означає, що зворотні відходи на стадії етикету дуже малі та стабільні. Тому варто приділяти увагу кваліфікації співробітника та мабуть додатково проаналізувати типи продукта та чи достатня в нас маржа.

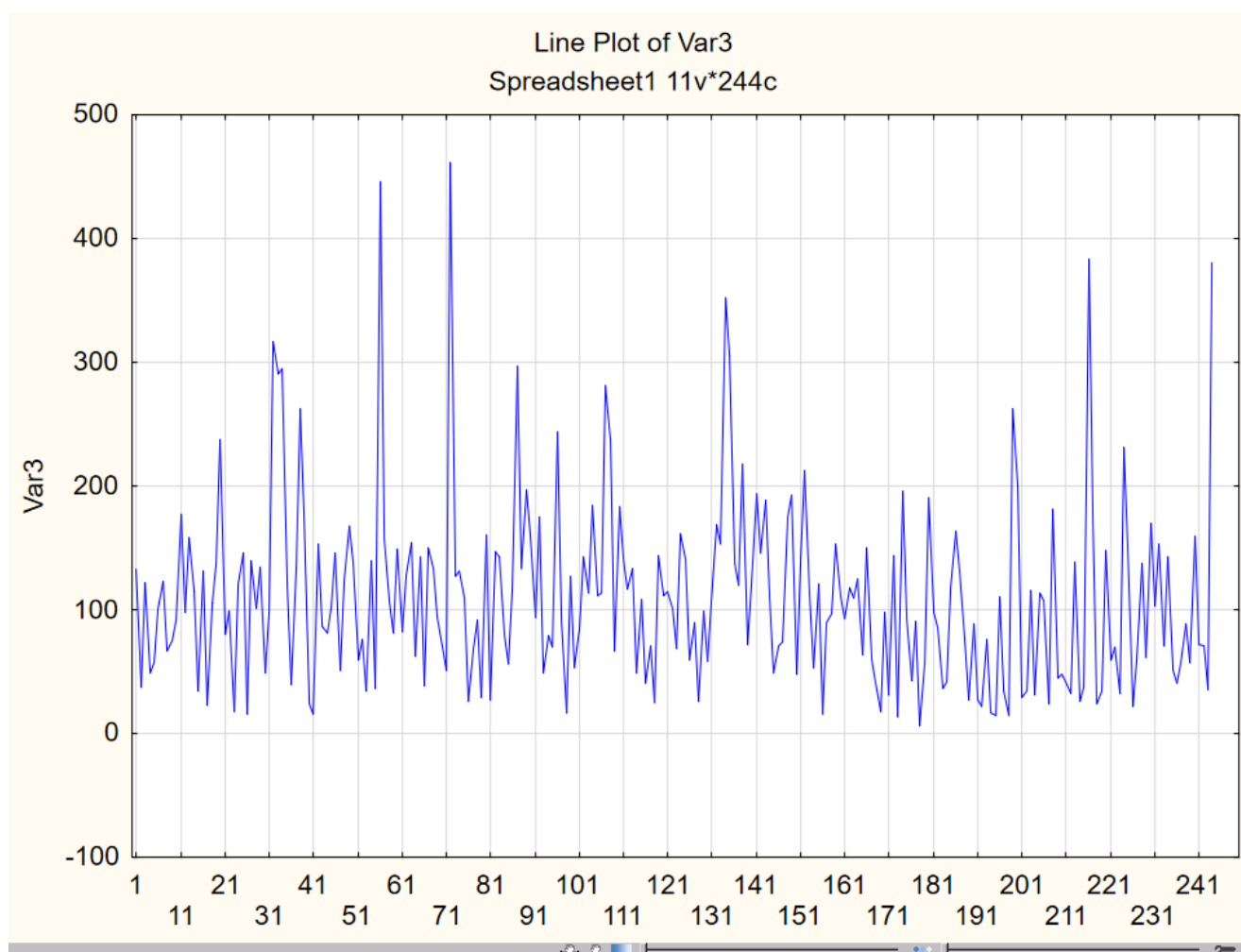


Рисунок 3.5.1 Залежність часу простою в залежності від зміни

Тепер перейдемо до наступної лінії – Winkler 2. У таблиці 3.5.2 наведено результати дтсперсійного аналізу

	SS	Степеней свободы	MS	F	p
Свободный член	4744328	1	4744238	494,3307	0,000000
Квалификация сотрудника	2684546	4	101256	5,0442	0,014553
Поставщик	1231546	3	25646	6,8566	0,0334654

етикета					
Тип продукта	5356	3	164589	24,3836	0,0498083
Количество остановок	456463	4	26574	2,1164	0,0431546

Таблиця 3.5.2 Результати аналізу лінії Winkler 2

Як бачимо, тут проблеми з етикетом набагато вагоміші ніж на попередній ліній, це через те що тут має місце слабка частина – загортка. А це у свою чергу через те, що загортальний автомат дуже чутливий до фотоміток. А оскільки різні постачальники мають різний колір то піднастраювати цю машину кожен раз виявляється дуже складно.

Приклад фотомітки на рисунку 3.5.2



Рисунок 3.5.2 Приклад фотоміток завдяки яким загортальний апарат розпізнає місце розрізу

Завдяки цьому аналізу було прийнято рішень купляти у більш дорожчого постачальника, для того щоб забезпечити безперервну роботу лінії. А з двома іншими постачальниками почалася певна праця по вирівнюванню кольорової гамми відміток.

Таким чином в нас є функція витрат у грошовому еквіваленті в залежності від того які зворотні відходи, скільки часу простояла лінія не випускаючи продукт, завдяки цьому можна оцінити які можна робити зміни та до чого вони призведуть.

Як ми бачимо, і на першій і на другій лінії дуже вагомий внесок це кваліфікація співробітника. Оскільки за допомогою кластерного аналізу ми поділили робітників на чотири категорії, а для кожної категорії коваріційний аналіз дав відповідь про середні значення зворотніх відходів, то можна оцінити скільки має коштувати навчання, для того щоб окупованість його була менше ніж півроку.

На основі статистичних даних отримали, що для лінії Winkler 1 це має бути менша сума ніж для лінії Winkler 2 через те що на лінії 1 більш дорожчі зворотні відходи, а отже і окупованість набагато швидше.

Висновки до розділу 3

У третьому розділі було описано та спроектовану систему підтримки прийняття рішень для аналізу та прийняття рішень щодо фінансових витрат на виробничій лінії. Дана система складається з таких структурних елементів: пристрої вводу-виводу, підсистема інтерфейсу користувача, база даних, підсистема обробки та аналізу інформації, блок прогнозування, і виведення результатів у вигляді графіків, таблиць, діаграм тощо.

На основі запропонованої СППР в рамках магістерської дисертації було розроблено програму для аналізу та впроваджено її у українську компанію. ПП дозволяє завантажувати дані, проводити попередній аналіз та обробку даних, будувати прогнозуючі моделі, а також обчислювати статистичні характеристики якості побудованої моделі та зберігати результати

прогнозування. Визначено мінімальні технічні характеристики персонального комп'ютера для коректної та повноцінної роботи програмного забезпечення, а саме: тактова частота процесору, об'єм оперативної пам'яті, об'єм пам'яті на диску, операційна система, додаткове програмне забезпечення, що підтримує роботу розробленого програмного продукту, та периферійні пристрої необхідні для повноцінної роботи оператора. Проведено детальний огляд інтерфейсу користувача. Розглянуто функціональні можливості програмного забезпечення та описану покрокову роботу з розробленою програмою з візуальним відображенням у вигляді рисунків робочого екрану програмного продукту. Розроблений програмний продукт апробовано на реальних статистичних даних з двох виробничих ліній які виробляють шоколад. Побудовано дві прогнозуючі моделі. Порівнявши статистичні характеристики побудованих моделей можна побачити, що ця програма дуже ефективна та оптимізує роботу аналітиків. Було проведено порівняльний аналіз методу логістичної регресії, з методами дерев рішень та логістичними, побудованими в системі Statistica. За критерієм загальної точності моделі найкращий результат виявився у методу дерев рішень.

РОЗДІЛ 4 СТАРТАП РОЗДІЛ

4.1 Опис ідеї проекту та її технологічний аудит

Опис ідеї стартап проекту наведено у таблиці 4.1

Таблиця 4.1 – Опис ідеї стартап проекту

Зміст ідеї	Напрямки застосування	Вигоди для користувача
Система підтримки прийняття рішень для оцінки та прогнозування роботи виробничої лінії	Надання інвесторам інформації про ймовірні фінансові ризики.	Інвесторам: допомога у прийнятті рішення щодо інвестування.
	Вибір напрямків фінансування	Гравцям на біржі: допомога при розробці стратегії торгівлі на біржі, набору позиції на ринку.
		Фінансовим аналітикам і консультантам: економія часу при аналізі потенційних позицій на ринку, порівняння отриманих даних щодо фін. Ризику, розрахованих на основі різних моделей.

Визначення сильних, слабких та нейтральних характеристик ідеї проекту продемонстровано у таблиці 4.2

Таблиця 4.2 – Визначення сильних, слабких та нейтральних характеристик ідеї проекту

№ п/ п	Техніко- економічні характеристики ідеї	(потенційні) товари/концепції конкурентів			W (сла бка стор она)	N (нейтрал ьна сторона)	S (силь на сторо на)
		Мій проект	prognosz.ru	Margincal. com			
1.	Ціна	Низька	Середня	Висока	+		
2.	Ефективність	Висока	Середня	Середня		+	
3.	Функціонал	Середній	Вузький	Широкий			+

Таблиця 4.3 – Технологічна здійсненність ідеї проекту

№ п/п	Ідея проекту	Технології її реалізації	Наявність технологій	Доступність технологій
1.	Розробка системи підтримки прийняття рішень для оцінки та прогнозування фінансових ризиків на фондовому ринку США.	Python	Наявна	Доступна
2.		MathLab	Наявна	Доступна
3.		MS Excel	Наявна	Доступна
Обрана технологія реалізації ідеї проекту: MS Excel				

4.2 Аналіз ринкових можливостей запуску стартап проекту

Таблиця 4.4 – Попередня характеристика потенційного ринку стартап проекту

№ п/ п	Показники стану ринку (найменування)	Характеристика
1.	Кількість головних гравців, од	2

2.	Загальний обсяг продаж, грн/ум.од	8500000
3.	Динаміка ринку (якісна оцінка)	Зростає

Продовження таблиці 4.4 – Попередня характеристика потенційного ринку стартап проекту

4.	Наявність обмежень для входу (вказати характер обмежень)	Важкодоступність певних даних
5.	Специфічні вимоги до стандартизації та сертифікації	Немає
6.	Середня норма рентабельності в галузі (або по ринку), %	30%

Таблиця 4.5 – Характеристика потенційних клієнтів стартап проекту

№ п/п	Потреба, що формує ринок	Цільова аудиторія (цільові сегменти ринку)	Відмінності у поведінці різних потенційних цільових груп клієнтів	Вимоги споживачів до товару
1.	Потреба у єдиному джерелі даних, що впливають на динаміку котирувань цінних паперів.	Інвестори	Зацікавлені у своєчасній і якісній аналітиці	Вимоги до своєчасності і достовірності даних, що публікуються
2.		Трейдери	Потребують об'єктивну оцінку потенційних фінансових ризиків для формування позицій на ринку	
3.		Фінансові аналітики і консультанти	Потребують своєчасну інформацію для проведення аналізу	

Таблиця 4.6 – Фактори загроз

№ п/п	Фактор	Зміст загрози	Можлива реакція компанії
1.	Новий продукт	Потенційні користувачі з підозрою ставляться	Поширення рекламної кампанії

		до нових продуктів	
--	--	--------------------	--

Таблиця 4.7 – Фактори можливостей

№ п/п	Фактор	Зміст можливості	Можлива реакція компанії
1.	Потреба у єдиному джерелі даних, що впливають на динаміку котирувань цінних паперів	Інвестори, трейдери та фінансові аналітики потребують ефективної платформи для аналізу потенційних можливостей і загроз на фондовому ринку	Задоволені інвестори, трейдери та фінансові аналітики

Таблиця 4.8 – Ступеневий аналіз конкуренції на ринку

Особливості конкурентного середовища	В чому проявляється дана характеристика	Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)
Тип конкуренції: олігополія	У сфері домінує невелика кількість подібних СППР	Поширення рекламної кампанії
За рівнем конкурентної боротьби: міжнаціональний	Наявна національна конкуренція	-
За галузевою ознакою: внутрішньогалузева	Наявна конкуренція в рамках одної галузі	-
Конкуренція за видами товарів: товарно-видова	Наявна конкуренція між схожими продуктами	Підвищення швидкості збору інформації, надання інструментів аналізу і прогнозування, збільшення кількості джерел інформації
За характером конкурентних переваг: нецінова	Наявна конкуренція завдяки розширенню бази даних та функціональних	Можливість вийти на ринок з недорогим продуктом

	можливостей	
За інтенсивністю: немарочна	Наявна конкуренція, де роль торгової марки незначна	-

Таблиця 4.9 – Аналіз конкуренції в галузі за М. Портером

	Прямі конкуренти в галузі	Потенційні конкуренти	Постачальники	Клієнти	Товари-замінники
Складові аналізу	prognosz.ru; Margincall.com	Доступ до каналів розподілу	Компанії на фондовому ринку, портали новин, банки	Швидкість надання інформації, її достовірність	Ціна
Висновки	Висока інтенсивність конкурентної боротьби з боку прямих конкурентів	Є можливість входу в ринок. Потенційних конкурентів немає.	Постачальники є джерелами інформації для СППР.	Клієнти вимагають своєчасності та достовірності інформації.	Аналогічні продукти конкурентів є дорогими, але їхній функціонал не є ширшим.

Таблиця 4.10 – Обґрунтування факторів конкурентоспроможності

№ п/п	Фактор конкурентоспроможності	Обґрунтування (наведення чинників, що роблять фактор для порівняння конкурентних проектів значущим)
1.	Ціна	Ціна запропонованого продукту нижча за ціни конкуруючих, але при цьому функціонал та якість інформації не поступається програмним продуктам конкурентів.

2.	Ефективність	За рахунок забезпечення своєчасності, повноти і достовірності інформації СППР дає можливість оцінити поточний стан на ринку, об'єктивно оцінити фінансові ризики та відкрити максимально вигідні позиції на фондовому ринку США.
Продовження таблиці 4.10		Обґрунтування факторів конкурентоспроможності

3.	Поріг входження	Так як у сфері не так багато аналогічний програмних продуктів, шанси увійти на ринок високі.
----	-----------------	--

Таблиця 4.11 – Порівняльний аналіз сильних та слабких сторін проекту

№ п/п	Фактор конкурентоспроможності	Бали 1-20	Рейтинг товарів-конкурентів у порівнянні з запропонованим продуктом						
			-3	-2	-1	0	+1	+2	+3
1.	Ціна	13	+						
2.	Ефективність	17						+	
3.	Поріг входження	18							+

Таблиця 4.12 – SWOT аналіз стартап проекту

<u>Сильні сторони:</u> Ефективність продукту Необхідність продукту Ціна продукту	<u>Слабкі сторони:</u> Невідомість продукту Вузьконаправленість продукту
<u>Можливості:</u> Охоплення широкої аудиторії зацікавлених користувачів – інвесторів, фінансових аналітиків, трейдерів. Підвищення ефективності роботи користувачів за рахунок розширення функціональних можливостей.	<u>Загрози:</u> Можлива незацікавленість продуктом через його специфічність і часту недовіру клієнтів до нових продуктів.

Таблиця 4.13 – Альтернативи ринкового впровадження стартап проекту

№ п/п	Альтернатива (орієнтовний комплекс заходів) ринкової	Ймовірність отримання	Строки реалізації
-------	--	-----------------------	-------------------

	поведінки	ресурсів	
1.	Розробка програмного забезпечення та грамотна маркетингова програма	Велика	4-5 місяців

4.3 Розробка ринкової стратегії проекту

Таблиця 4.14 – Вибір цільових груп потенційних клієнтів

№ п/п	Опис профілю цільової групи потенційних клієнтів	Готовність споживачів сприйняти продукт	Орієнтовний попит в межах цільової групи (сегменту)	Інтенсивність конкуренції в сегменті	Простота входу у сегмент
1.	Інвестори	Значна готовність	Високий	Низька	Середня
2.	Трейдери	Значна готовність	Високий	Низька	Висока
3.	Фінансові аналітики і консультанти	Значна готовність	Високий	Низька	Висока
Які цільові групи обрано: Інвестори, трейдери, фінансові консультанти та аналітики					

Таблиця 4.15 – Визначення базової стратегії розвитку

№ п/п	Обрана альтернатива розвитку проекту	Стратегія охоплення ринку	Ключові конкурентоспроможні позиції відповідно до	Базова стратегія розвитку
-------	--------------------------------------	---------------------------	---	---------------------------

			обраної альтернативи	
1.	Розробка програмного забезпечення та грамотна маркетингова програма	За рахунок потреби в широкофункціональному продукті	Ефективність	Стратегія спеціалізації

Таблиця 4.16 – Визначення базової стратегії конкурентної поведінки

№ п/п	Чи є проект «першопрохідцем» на ринку?	Чи буде компанія шукати нових споживачів, або забирати існуючих у конкурентів?	Чи буде компанія копіювати основні характеристики товару конкурента, і які?	Стратегія конкурентної поведінки
1.	Ні	Шукати нових споживачів і забирати існуючих у конкурентів	Буде розроблений продукт із ширшим функціоналом і повнішою інформаційною базою	Стратегія заняття конкурентної ніші.

Таблиця 4.17 – Визначення стратегії позиціонування

№ п/п	Вимоги до товару цільової аудиторії	Базова стратегія розвитку	Ключові конкурентоспроможні позиції власного стартап-проекту	Вибір асоціацій, які мають сформулювати комплексну позицію власного проекту (три ключових)
1.	Необхідність СППР з повною, достовірною і своєчасною інформацією та	Стратегія спеціалізації	Ефективність	Висока ефективність Простота у використанні

	інструментарієм для аналізу ринку			
--	---	--	--	--

4.4 Розробка маркетингової програми стартап проекту

Таблиця 4.18 – Визначення ключових переваг концепції потенційного товару

№ п/п	Потреба	Вигода, яку пропонує товар	Ключові переваги перед конкурентами (існуючі або такі, що потрібно створити)
1.	Потреба в ефективному продукті	Пропонує ефективний продукт	Ефективність товару вище ніж ефективність товару конкурентів

Таблиця 4.19 – Опис трьох рівнів моделі товару

Рівні товару	Сутність та складові		
I. Товар за задумом	Система підтримки прийняття рішень для оцінки та прогнозування фінансових ризиків на основі різних моделей та підходів на фондовому ринку за США.		
II. Товар у реальному виконанні	Властивості/характеристики	М/Нм	Вр/Тх /Тл/Е/Ор
	1. Мультиплатформенність 2. Зручний інтуїтивний інтерейс	-	-
	Якість: стандарти ефективності		
	Пакування: електронне розповсюдження		
	Марка: JMarketRisk		
III. Товар із підкріпленням	До продажу: -		
	Після продажу: технічна підтримка		
За рахунок чого потенційний товар буде захищено від копіювання: захист інтелектуальної власності			

Таблиця 4.20 – Визначення меж встановлення ціни

№ п/п	Рівень цін на товари-замінники	Рівень цін на товари-аналоги	Рівень доходів цільової групи споживачів	Верхня та нижня межі встановлення ціни на товар/послугу
1.	-	9500000-63000000 грн.	30000+ грн.	10000-20000 грн.

Таблиця 4.21 – Формування системи збуту

№ п/п	Специфіка закупівельної поведінки цільових клієнтів	Функції збуту, які має виконувати постачальник товару	Глибина каналу збуту	Оптимальна система збуту
1.	Закупівля через інтернет	Підтримка нормального функціонування сайту	0	Електронне розповсюдження

Таблиця 4.22 – Концепція маркетингових комунікацій

№ п/п	Специфіка поведінки цільових клієнтів	Канали комунікацій, якими користуються цільові клієнти	Ключові позиції, обрані для позиціонування	Завдання рекламного повідомлення	Концепція рекламного звернення
1.	Потреба в ефективному продукті	Інтернет-мережі	Висока ефективність Простота у використанні	Провести якісну маркетингову кампанію	Донести специфіку продукту

4.5 Висновки

Отже, існує можливість ринкової комерціалізації проекту, так як існує попит, наявна динаміка ринку та рентабельність роботи на маркеті.

Звичайно, існують перспективи впровадження проекту з огляду на потенційні групи клієнтів, бар'єри входження, поточну конкуренцію та конкурентноспроможність стартапу. Раціонально обрати альтернативу розробку програмного забезпечення та грамотну маркетингову програму для ринкової реалізації проекту. Вважаємо, що подальша імплементація проекту доцільна.

ВИСНОВКИ

Дана робота присвячена аналізу, моделюванню та прогнозуванню економетричних показників із використанням Statistica, мови програмування R, а також за допомогою методів математичних моделей у вигляді дисперсійного, коваріаційного та лінійно-імовірносного аналізів. Для обчислення критеріїв якості обраних моделей було використано бібліотеки мови R.

Перший розділ дипломної роботи присвячено опису фінансово - виробничої інформації щодо економетричних даних та існуючих методів щодо їх аналізу.

В другому розділі наведено детальний опис математичних моделей, що будуть використані в подальшому, описано основні методи, процедури і функції для аналізу, моделювання та прогнозування часових рядів.

В третьому розділі наведена детальна теоретична інформація щодо методу розв'язання задач для вибору стратегій виробництва. Також запропонована структура системи підтримки прийняття рішень, моделювання та прогнозування економетричних даних.

В четвертому розділі представлений економічний аналіз зробленої роботи. Як кінцевий результат, отримано варіант реалізації програмного продукту, що передбачає зручний користувацький інтерфейс, повноцінний і якісний функціонал для створення і редагування вибору стратегії виробництва і високий рівень швидкодії.

Подальшими напрямками роботи можуть бути питання, що стосуються:

1. Розробити повноцінний корпоративний додаток, з різними правами доступу, для вибору стратегії виробництва, а також відслідкування процесу виробництва. З можливим редагуванням обраної стратегії.

2. Застосування інших методів та моделей для аналізу та прогнозування економетричних показників. Наприклад нейронні мережі, дерева рішень для неперервних даних, методи дисперсійної та коваріаційного аналізів, та інших методів інтелектуального аналізу даних (ІАД).

3. Удосконалення моделі, розробляючи її для більш складних ліній. За рахунок внесення нових обмежень та можливостей.

Перелік посилань

1. Bielecki T. R. Credit Risk: Modeling, Valuation, Hedging / T. R. Bielecki, M. Rutkowski. – Berlin: Springer, 2002. – 500 p.
2. Van Gruening H. Analyzing and Managing Banking Risks / H. Van Gruening, S. B. Bratanovic. – Washington: The World Bank, 2003. – 386 p.
3. Aven T. Foundations of Risk Analysis: A Knowledge and Decision-Oriented Perspective / T. Aven. – New York: John Wiley & Sons, 2003. – 198 p.
4. Муравйова М. Ю. Шляхи вдосконалення оцінки кредитоспроможності позичальників банками України [Електронний ресурс] / М. Ю. Муравйова // Управління розвитком. – 2012. – №14 (135). – Режим доступу:http://www.nbu.gov.ua/old_jrn/Soc_Gum/Uproz/2012_14/u1214_mur.pdf.
5. Вдовенко Л. О. Економічна сутність та значення кредитоспроможності підприємств [Електронний ресурс] / Л. О. Вдовенко // Облік і фінанси. – 2012. – № 1. – С. 108-111. – Режим доступу: http://nbuv.gov.ua/UJRN/Oif_apk_2012_1_23.
6. Бучко І. Є. Скоринг як метод зниження кредитного ризику банку [Електронний ресурс] / І. Є. Бучко // Вісник Університету банківської справи Національного банку України. – 2013. – № 2. – Режим доступу:http://nbuv.gov.ua/UJRN/VUbsNbU_2013_2_37.
7. Клейнер Г. Б. История современного кредитного скоринга [Электронный ресурс] / Г. Б. Клейнер, Д. С. Коробов // Проблемы региональной экономики: электронный журнал. – 2012. – Вып. 17. – С. 45-49. Режим доступа: <http://www.regec.ru/articles/2012/vol1/5.pdf>.
8. Ишина И. В. Скоринг - модель оценки кредитного риска [Электронный ресурс] / И. В. Ишина, М. Н. Сазонова // Аудит и финансовый анализ. – 2007. – № 4. – Режим доступа: <http://www.auditfin.com/fin/2007/4/Ishina/Ishina%20.pdf>.
9. Самойлова С. С. Скоринговые модели оценки кредитного риска [Электронный ресурс] / С. С. Самойлова, М. А. Курочка //

Социальноэкономические явления и процессы. – 2014. – № 3 (61). – Режим доступа: <http://cyberleninka.ru/article/n/skoringovye-modeli-otsenikreditnogo-riska>.

10. Лункіна Т. І. Використання скоринг моделі при управлінні ризиками споживчого кредитування [Електронний ресурс] / Т. І. Лугкіна // Ефективна економіка. – 2015. – № 2. – Режим доступу: <http://www.economy.nayka.com.ua/?op=1&z=3792>.

11. Терентьев А. Н. SAS BASE: Основы программирования / А. Н. Терентьев, В. Н. Домрачев, Р. И. Костецкий – К.: Эдельвейс, 2014. – 304 с.

12. Anderson B. S. Developing Credit Scorecards Using SAS Credit Scoring for Enterprise Miner 5.3 / B. S. Anderson, R. W. Thompson. – Cary: SAS Institute Inc, 2009. – 41 p.

13. Кожухівська О. А. Прогнозування ризиків кредитування фізичних осіб за математичними моделями [Електронний ресурс] / О. А. Кожухівська // Вісник Національного університету «Львівська політехніка». Інформаційні системи та мережі. – 2013. – № 770. – С. 177-185. – Режим доступу: http://nbuv.gov.ua/UJRN/VNULPICM_2013_770_23.

14. Романенко А. В. Логистическая регрессия // Молодёжь и наука: Сборник материалов VIII Всероссийской научно-технической конференции студентов, аспирантов и молодых учёных, посвященной 155-летию со дня рождения К. Э. Циолковского [Электронный ресурс]. – Красноярск: Сибирский федеральный ун-т, 2012. – Режим доступа: <http://conf.sfu-kras.ru/sites/mn2012/section21.html>

15. Калиткин Н. Н. Численные методы / Н.Н. Калиткин. – М.: Наука, 1978. – 592 с.

16. Сиддики Н. Скоринговые карты для оценки кредитных рисков. Разработка и внедрение интеллектуальных методов кредитного скоринга / Н. Сиддики; пер. с англ. Е. Ильичева. – М.: Манн, Иванов и Фербер, 2014. – 268 с.

17. Сорокин А. С. Построение скоринговых карт с использованием модели логистической регрессии [Электронный ресурс] / А. С. Сорокин // Наукovedение. – 2014. – Вып. 2 (21). – Режим доступа: <http://naukovedenie.ru/PDF/180EVN214.pdf>.

18. Кузнєцова Н. В. Порівняльний аналіз характеристик моделей оцінювання ризиків кредитування / Н. В. Кузнєцова, П. І. Бідюк // Наукові вісті НТУУ «КПІ». – 2010. – №1. – С. 42-53.

19. Сорокин А.С. К вопросу валидации модели логистической регрессии в кредитном скоринге [Электронный ресурс] / А. С. Сорокин // Наукovedение. – 2014. – Вып. 2 (21). – Режим доступа: <http://naukovedenie.ru/PDF/173EVN214.pdf>.